PTO-1542
(4-85)

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE

ONLINE SEARCH REQUEST FORM

USER Spector U.S. SERIAL NUMBER 243545

ART UNIT 1812 PHONE 308-1793 DATE 4/3/95

Please give a detailed statement of requirements. Describe as specifically as possible the subject matter to be searched. Define any terms that may have special meaning. Give examples or relevant citations, authors, or keywords, if known.

You may include a copy of the broadest and or relevant claim(s).

⇒ See CRF for parent 08/162407

Please search Seq ID 145,

also ~~a~~ ~~b~~ reverse translate

& search aa 28-182 of ID6.

Thanks!

STAFF USE ONLY

4-23

COMPLETED 4/5

SEARCHER Linda 4058

ONLINE TIME 32 (in minutes) TOTAL TIME

NO. OF DATABASES

SYSTEMS
- CAS ONLINE
- DARC/QUESTEL
- DIALOG
- SDC
- ✓ OTHER JG

```
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq1.res made by  on Wed 5 Apr 95 19:41:35-PDT.


Query sequence being compared:US-08-162-407-1 (1-879)
Number of sequences searched:               313646
Number of scores above cutoff:                4589

      Results of the initial comparison of US-08-162-407-1 (1-879) with:
   Data bank : EMBL-NEW 1, all entries
   Data bank : GenBank 86, all entries
   Data bank : GenBank-NEW 1, all entries
   Data bank : UEMBL 41_86, all entries

1000000-
       -
N      -
U00000-
M      -
B      -
E      -
R      -
       -
O      -**
F00000-
       -
S      -  *
E50000-
Q      -
U      -
E      -
N      -   *
C      *
E      -
S10000-
       -
       -    *
  5000-
       -
       -
       -
       -     *
       -
       -
  1000-
       -
       -      *
   500-
       -
       -
       -
       -       *
       -
       -
```


Spector 243543/162407

```
100-        *
  -
  -
 50-
  -
  -
  -         *
  -          *
  -
  -
 10-
  -
  -
  5-           *
  -
  -            *
  -                        *
  -
  -
  -                                                  *                *
  0-------------------------------------------------------------------------
    ||  |||||  || |      |      |      |      |      |      |      |      |
SCORE 0|  ||92|  |184    276    368    461    553    645    737    829
STDEV -1      6   9
```

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	31	30	13.25

Times:	CPU	Total Elapsed
	00:38:23.03	00:38:32.00

```
Number of residues:              288682119
Number of sequences searched:       313646
Number of scores above cutoff:        4589

Cut-off raised to 23.
Cut-off raised to 27.
Cut-off raised to 30.
Cut-off raised to 33.
Cut-off raised to 36.
Cut-off raised to 39.
Cut-off raised to 42.
Cut-off raised to 45.
Cut-off raised to 48.
Cut-off raised to 52.
```

Cut-off raised to 55.
Cut-off raised to 58.
Cut-off raised to 61.
Cut-off raised to 64.
Cut-off raised to 67.
Cut-off raised to 70.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 60 standard deviations above mean ****					
1. MUSLIGAND	Mouse flt3 ligand mRNA, compl	829	829	829	60.20	0
	**** 45 standard deviations above mean ****					
2. MMU04807	Mus musculus FLT3/FLK2 ligand	1152	630	807	45.19	0
	**** 26 standard deviations above mean ****					
3. HSU04806	Human FLT3/FLK2 ligand mRNA,	859	378	591	26.18	0
4. HSU03858	Human flt3 ligand mRNA, compl	1080	378	608	26.18	0
	**** 10 standard deviations above mean ****					
5. HSC1RCP	Human complement C1r mRNA, co	2493	173	397	10.71	0
6. HUMIRBPG	Human interstitial retinol-bi	9711	172	391	10.64	0
7. HUMIRBPG	Human interstitial retinol-bi	9711	172	391	10.64	0
	**** 9 standard deviations above mean ****					
8. HSCGPDEG2	Human rod cGMP-phosphodiester	2417	162	392	9.88	0
9. HSC1R	Human mRNA for complement com	2386	160	398	9.73	0
10. HUMC1R	Human complement C1r mRNA, co	2493	160	397	9.73	0
11. HUMCEL	Human carboxyl ester lipase (	11502	156	395	9.43	0
12. HSCEL	Human carboxyl ester lipase (	11502	156	395	9.43	0
13. DDU03413	Dictyostelium discoideum AX2	1679	155	369	9.36	0
14. HSNUMAU4G	H.sapiens NuMA gene (Clone U4	2832	153	403	9.20	0
15. HSNUMAU6G	H.sapiens NuMA gene (Clone U6	4762	153	403	9.20	0
16. HSNUMAMR	H.sapiens mRNA for NuMA prote	7154	153	403	9.20	0
17. SSENDOGA	S.scrofa endoglin mRNA, compl	2908	151	390	9.05	0
	**** 8 standard deviations above mean ****					
18. OCAPOAI	Rabbit apo-AI mRNA for apolip	892	149	330	8.90	0
19. MUSACTR2B	Mouse activin receptor (ActR	1708	148	389	8.83	0
20. LEMINVOL	Lemur (ring-tailed) involucri	1380	147	396	8.75	0
21. HUMIRBPM	Human interstitial retinol bi	4124	146	352	8.68	0
22. HUMIRBPM	Human interstitial retinol bi	4124	146	352	8.68	0
23. MUSTRANONC	Mus musculus ERP mRNA, comple	2211	145	400	8.60	0
24. AD4POLTP2	Adenovirus type 4 Pol gene, e	5541	145	386	8.60	0
25. HUMMEVKIN	Homo sapiens mevalonate kinas	1967	144	407	8.53	0
26. MMNETRN	M.musculus net mRNA.	2658	144	398	8.53	0
27. SCSECYDNA	S.coelicolor secY locus DNA.	4794	143	411	8.45	0
28. SCSECYDNA	S.coelicolor secY locus DNA	4794	143	411	8.45	0
29. HSAE1ERY	Human anion exchanger (AE1) g	21319	140	396	8.22	0
30. AGMUVDAMBP	Cercopithecus aethiops UV-dam	4181	138	382	8.07	0
	**** 7 standard deviations above mean ****					
31. HRSIFNB	Horse interferon-beta gene, c	2467	135	349	7.85	0
32. S70756S3	Tlx-1=homeobox gene HOX11 hom	771	134	339	7.77	0
33. OCCALPH13	O.cuniculus gene for C alpha1	1422	134	393	7.77	0
34. CRATP2	C.reinhardtii atp2 (atpB) mRN	2664	134	404	7.77	0
35. MUSACMHCB	Mouse alpha cardiac myosin he	5817	133	383	7.70	0

36. I08072	Sequence 4 from patent EP 029	1698	132	407	7.62	0
37. PIGA2AR	Porcine alpha2A-adrenergic re	1728	132	397	7.62	0
38. HUMGLUT5	Human glucose transport-like	2218	132	395	7.62	0
39. HUMGCK	Human glucokinase (GCK) gene,	7807	132	385	7.62	0
40. HUMIHRP	Human placental ribonuclease	1682	131	407	7.54	0
41. HUMIHRP	Human placental ribonuclease	1682	131	407	7.54	0
42. HUMRAI	Human ribonuclease/angiogenin	1920	131	407	7.54	0
43. HUMRAI	Human ribonuclease/angiogenin	1920	131	407	7.54	0
44. HSRAI	Human mRNA for ribonuclease/a	1921	131	407	7.54	0
45. NEUAMG	N.crassa am (NADP-specific gl	2643	131	389	7.54	0

1. US-08-162-407-1 (1-879)
   MUSLIGAND Mouse flt3 ligand mRNA, complete cds.

```
LOCUS       MUSLIGAND     829 bp ss-mRNA            ROD       04-JAN-1994
DEFINITION  Mouse flt3 ligand mRNA, complete cds.
ACCESSION   L23636
KEYWORDS    ligand.
SOURCE      Mus musculus (strain SJL/J) cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
REFERENCE   1  (bases 1 to 829)
  AUTHORS   Lyman,S.D., James,L., VandenBos,T., deVries,P., Brasel,K.,
            Gliniak,B., Hollingsworth,L., Picha,K.S., McKenna,H.J.,
            Splett,R.R., Fletcher,F.A., Maraskovsky,E., Farrah,T.,
            Foxworthe,D., Williams,D.E. and Beckmann,M.P.
  TITLE     Molecular cloning of a ligand for the flt3/flk-2 tyrosine kinase
            receptor -- a proliferative factor for primitive hematopoietic
            cells
  JOURNAL   Cell 75, 1157-1167 (1993)
  STANDARD  full automatic
COMMENT     NCBI gi: 439441
FEATURES             Location/Qualifiers
     source          1..829
                     /organism="Mus musculus"
                     /strain="SJL/J"
                     /sequenced_mol="cDNA to mRNA"
     sig_peptide     32..109
                     /codon_start=1
     CDS             32..727
                     /note="NCBI gi: 439442"
                     /codon_start=1
                     /product="flt3 ligand"
                     /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                     KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                     NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLALKPCIGKACQNFSRCLEVQ
                     CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLPLTLVLLAAAWGLRWQRARRR
                     GELHPGVPLPSHP"
     mat_peptide     110..724
                     /codon_start=1
BASE COUNT      163 a    265 c    221 g    180 t
ORIGIN
```

Initial Score = 829 Optimized Score = 829 Significance = 60.20
Residue Identity = 100% Matches = 829 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

20 X 30 40 50 60 70 80

```
ACGACCTGCTCTGTCACAGGCATGAGGGGTCCCCGGCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAA
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           CTGTCACAGGCATGAGGGGTCCCCGGCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAA
           X       10        20        30        40        50        60

  90       100       110       120       130       140       150
ATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGTCCTTGCCTGCGGGGGACACCTGACTGTTACTTCAGCC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGTCCTTGCCTGCGGGGGACACCTGACTGTTACTTCAGCC
      70        80        90        100       110       120       130

160       170       180       190       200       210       220       230
ACAGTCCCATCTCCTCCAACTTCAAAGTGAAGTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACAGTCCCATCTCCTCCAACTTCAAAGTGAAGTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAG
   140       150       160       170       180       190       200

       240       250       260       270       280       290       300
TCACTGTGGCCGTCAATCTTCAGGACGAGAAGCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TCACTGTGGCCGTCAATCTTCAGGACGAGAAGCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCT
 210       220       230       240       250       260       270

     310       320       330       340       350       360       370
GGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATAC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATAC
280       290       300       310       320       330       340       350

  380       390       400       410       420       430       440
ATTTTGTCACCTCATGTACCTTCCAGCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATTTTGTCACCTCATGTACCTTCCAGCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACC
      360       370       380       390       400       410       420

450       460       470       480       490       500       510
TCCTGAAGGACACCTGCACACAGCTGCTTGCTCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TCCTGAAGGACACCTGCACACAGCTGCTTGCTCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTC
     430       440       450       460       470       480       490

520       530       540       550       560       570       580       590
GGTGCCTGGAGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GGTGCCTGGAGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAG
   500       510       520       530       540       550       560

       600       610       620       630       640       650       660
CCACGGAGCTCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACTGGTGC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CCACGGAGCTCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACTGGTGC
 570       580       590       600       610       620       630

     670       680       690       700       710       720       730
TGCTGGCAGCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGGGGTGCCCC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TGCTGGCAGCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGGGGTGCCCC
640       650       660       670       680       690       700       710

   740       750       760       770       780       790       800
```



```
TCCCCTCCCATCCCTAGGATTCGAGCCTTGTGCATCGTTGACTCAGCCAGGGTCTTATCTCGGTTACACCTG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TCCCCTCCCATCCCTAGGATTCGAGCCTTGTGCATCGTTGACTCAGCCAGGGTCTTATCTCGGTTACACCTG
       720       730       740       750       760       770       780

810       820       830       840       850   X   860
TAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGAGCAGGTCGT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGG
     790       800       810       820       X
```

2. US-08-162-407-1 (1-879)
   MMU04807 Mus musculus FLT3/FLK2 ligand mRNA, complete cds.

```
LOCUS       MMU04807     1152 bp    mRNA            ROD       11-MAY-1994
DEFINITION  Mus musculus FLT3/FLK2 ligand mRNA, complete cds.
ACCESSION   U04807
KEYWORDS    .
SOURCE      mouse.
  ORGANISM  Mus musculus
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Rodentia; Myomorpha; Muridae; Mus.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
            Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
            Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
            Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
            Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
  TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
            haematopoietic stem cells and is encoded by variant RNAs
  JOURNAL   Nature 368, 643-648 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1152)
  AUTHORS   Culpepper,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
            Institute, 901 California Avenue, Palo Alto, CA 94304, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 483846
FEATURES             Location/Qualifiers
     source          1..1152
                     /clone="mouse Flt3/Flk2 ligand T110"
                     /clone_lib="TA4 cDNA"
                     /organism="Mus musculus"
                     /cell_line="TA4 stroma"
     CDS             256..954
                     /note="NCBI gi: 483847"
                     /codon_start=1
                     /product="FLT3/FLK2 ligand"
                     /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                     KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                     NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLGLKPCIGKACQNFSRCLEVQ
                     CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLLPLTLVLLAAAWGLRWQRARR
                     RGELHPGVPLPSHP"
BASE COUNT      238 a    346 c    324 g    244 t
ORIGIN

Initial Score    =    630  Optimized Score  =    807  Significance = 45.19
Residue Identity =    90%  Matches          =    828  Mismatches   =    48
```

```
Gaps            =     35  Conservative Substitutions          =     0

            X         10        20                  30        40
            GT-CGA-CTGGA-ACGAGACGACCTG---CT----------CTGTCACAGGCATGAGGGGTCC
            ||  || |  || ||  | |  ||||   ||          ||||||||||||||||||||||
    GGGCTGCAGGGTATGAGCCCGAGACCTGCCCTCCTGTCACTTCCAAGAACCTGTCACAGGCATGAGGGGTCC
       180    X 190       200       210       220       230       240

     50        60        70        80        90       100       110
    CCGGCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTG
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    CCGGCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTG
     250       260       270       280       290       300       310

  120       130       140       150       160       170       180       190
    AGTCCTTGCCTGCGGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAG
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    AGTCCTTGCCTGCGGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAG
  320       330       340       350       360       370       380       390

         200       210       220       230       240       250       260
    TTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAGAAG
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    TTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAGAAG
           400       410       420       430       440       450       460

       270       280       290       300       310       320       330
    CACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCT
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    CACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCT
         470       480       490       500       510       520       530

     340       350       360       370       380       390       400
    AAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCTA
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    AAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCTA
       540       550       560       570       580       590       600

    410       420       430       440       450       460       470
    CCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCTGCTTGCT
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
    CCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCTGCTTGGT
     610       620       630       640       650       660       670

  480       490       500       510       520       530       540       550
    CTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTCC
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    CTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTCC
  680       690       700       710       720       730       740       750

         560       570       580       590       600       610       620
    TCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCAG
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    TCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCAG
           760       770       780       790       800       810       820

       630           640       650       660       670       680       690
    CTGTTGCTC---CTGCTGCTGCTGCTGCCTCTCACACTGGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTGG
    |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    CTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACTGGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTGG
```



```
         830       840       850       860       870       880       890

         700       710       720       730       740       750       760
    CAAAGGGCAAGAAGGAGGGGGGGAGCTCCACCCTGGGGTGCCCCTCCCCTCCCATCCCTAGGATTCGAGCCTT
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
    CAAAGGGCAAGAAGGAGGGGGGGAGCTCCACCCTGGGGTGCCCCTCCCCTCCCATCCCTAGGATGCGAGCCTT
       900       910       920       930       940       950       960

       770       780       790       800       810                 820
    GTGCATCGTTGACTCAGCCAGGGTCTTATCTCGGTTACACCTGTA-AT-----CTCA----GCCCTTGGGAGC
    |||||||||||||||||||||||||||||| | |   || | ||    | ||    | | ||| | |
    GTGCATCGTTGACTCAGCCAGGGTCTTATCTCGATGAGGTCTCAATATGTTGCCCAAACTGACTTTGAAAAC
     970       980       990      1000      1010      1020      1030

    830           840       850       860       870        X
    CCAGA-GCAGGAT--TGCTGAATGGTCTGGAGCAGGTCGTCTCG--TTCCAGTCGAC
    |  || |||   |  |||   |   ||   ||   |  || |  || | | |
    CTCGATGCACCTTCCTGCCCCACAAACT--TCCA-AACAGCTGGGCTTACGGGCATGCTATATACAA
 1040      1050      1060      1070      1080      1090 X    1100
```

3. US-08-162-407-1 (1-879)
   HSU04806 Human FLT3/FLK2 ligand mRNA, complete cds.

```
LOCUS       HSU04806      859 bp    mRNA            PRI       11-MAY-1994
DEFINITION  Human FLT3/FLK2 ligand mRNA, complete cds.
ACCESSION   U04806
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
            Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
            Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
            Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
            Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
  TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
            haematopoietic stem cells and is encoded by variant RNAs
  JOURNAL   Nature 368, 643-648 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 859)
  AUTHORS   Culpepper,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
            Institute, 901 California Avenue, Palo Alto, CA 94304, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 483844
FEATURES             Location/Qualifiers
     source          1..859
                     /clone="human Flt3/Flk2 ligand S86"
                     /clone_lib="SV48 cDNA"
                     /organism="Homo sapiens"
                     /cell_line="SV48 thymic stroma"
     CDS             93..800
                     /note="NCBI gi: 483845"
                     /codon_start=1
                     /product="FLT3/FLK2 ligand"
                     /translation="MTVLAPAWSPTTYLLLLLLLSSGLSGTQDCSFQHSPISSDFAVK
```

Same as 2

```
                    IRELSDYLLQDYPVTVASNLQDEELCGALWRLVLAQRWMERLKTVAGSKMQGLLERVN
                    TEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQP
                    DSSTLPPPWSPRPLEATAPTAPQPPLLLLLLLPVGLLLLAAAWCLHWQRTRRRTPRPG
                    EQVPPVPSPQDLLLVEH"
 BASE COUNT      147 a    318 c    246 g    148 t
 ORIGIN

Initial Score    =    378  Optimized Score  =    591  Significance = 26.18
Residue Identity =    72%  Matches          =    613  Mismatches   =   200
Gaps             =     32  Conservative Substitutions               =     0

            X       10        20        30        40        50        60
            GTCGACTGGAACGAGAC-GACCTGCTCTGTCA-CAGGCATGAGGGGTCCCCGGCAGAGATGA
              | |||||  | || | ||||  |   | ||  ||||||||||| ||||||| || ||||
  CCCTTCCACACCCAACTGGGGCAAGCCTGACC--CGGCGACAGGAGGCATGAGGGGCCCCCGGCCGAAATGA
    30      X 40        50        60        70        80        90

          70        80        90       100       110       120       130
  CAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGTCCTTGCCTGC
  ||||||||||||||||||||||||||||    |||  |   | || ||| ||||||||||||  |  | ||
  CAGTGCTGGCGCCAGCCTGGAGCCCAACAACCTATC---TCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCA
   100       110       120       130       140       150       160

       140       150       160       170       180       190       200
  GGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAGTTTAGAGAGTTGA
  | ||||| |  |||||  | ||||   ||||| ||||||||||| |||||   || ||  |  | ||| ||
  GTGGGACCCAGGACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGT
    170       180       190       200       210       220       230

     210       220       230       240       250       260       270
  CTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAGAAGCACTGCAAGGCCT
  |||||  |||||||  ||||||||||||||||  ||||||  ||| || |||||||| ||| ||||  |||
  CTGACTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGCGC
  240       250       260       270       280       290       300

    280       290       300       310       320       330       340
  TGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATGCAAACGC
  | ||| | ||  |||| ||  |||||||||||| ||||  || ||||||| || ||||| ||||||| || ||
  TCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGC-AAGGC
310       320       330       340       350       360       370       380

  350       360       370       380       390       400       410
  TT-CTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCTACCAGAATGTCTG
  || ||||||  ||| |||||  |||||||| |||||||||  |||| |||| |||||||  ||    |||||
  TTGCTGGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTT
         390       400       410       420       430       440       450

420       430       440       450       460       470       480       490
  CGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCTGCTTGCTCTGAAGCCCTGT
  || ||||||||||||||||||||||  |||||||  |||| |||| |   |||||| | || ||||||||||
  CGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGG
       460       470       480       490       500       510       520

        500       510       520       530       540       550       560
  ATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCTG
  |||    |  |  |||||||  |||||  |||||||||||  |||||||  |||||  ||||||||  |||||||
  ATC----ACTC--GCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCA
         530         540       550       560       570       580       590

      570       580       590       600       610       620       630
```



```
  CCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTG
  |||||| ||||||||    |||| || |||||  ||| || | |||| ||   ||| |  ||  |||||||
  CCCCCATGGAGTCCCCGGCCCCTGGAGGCCAC--AGCCCCGACAGCCCCG---CAGCCCCCT-CTGCTCCTC
         600       610       620       630       640       650

    640       650       660       670       680       690       700
  CTGCTGCTGCTGCCTCTCACACTGGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGG
  || ||||||||||  |    ||  |||||||||| || |||||| |||| | |||||| ||| |  |  ||
  CTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCGG
   660       670       680       690       700       710       720

  710       720       730       740       750       760       770
  AGGGGGGAGCTCCACCCTGGGGTGC----CCCTCCCCTCCC--ATCCCTAGGATTCGAGCCTTGTGCATCGT
  |  ||  | | || |||||||| ||     || ||| ||||   |||| ||||   |   |||||| | |
  A--GGACACC-CCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCAC
   730       740       750       760       770       780       790

      780       790       800       810       820       830     X  840
  TGACTCAGCCAGGGTCTTAT-CTCGGTTACACCTGTAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTG
  ||||   |||| || || || || ||    |   |||  | || |    |||||  |   |||
  TGACCTGGCCAAGGCCTCATCCTGGGGAGGATACGTAGGCACA-CAGAGGGGAGTCACCAGCC
  800       810       820       830       840       850       X

  A
```

4. US-08-162-407-1 (1-879)
   HSU03858 Human flt3 ligand mRNA, complete cds.

```
 LOCUS       HSU03858     1080 bp    mRNA            PRI       19-JUL-1994
 DEFINITION  Human flt3 ligand mRNA, complete cds.
 ACCESSION   U03858
 KEYWORDS    .
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 1080)
   AUTHORS   Lyman,S.D., James,L., Johnson,L., Brasel,K., de Vries,P.,
             Escobar,S.S., Downey,H., Splett,R.R., Beckmann,M.P. and
             McKenna,H.J.
   TITLE     Cloning of the human homologue of the murine flt3 ligand: a growth
             factor for early hematopoietic progenitor cells
   JOURNAL   Blood 83, 2795-2801 (1994)
   STANDARD  full automatic
 REFERENCE   2  (bases 1 to 1080)
   AUTHORS   Lyman,S.D.
   TITLE     Direct Submission
   JOURNAL   Submitted (30-NOV-1993) Stewart D. Lyman, Immunex Research and
             Development Corporation, 51, University St., Seattle, WA 98101, USA
   STANDARD  full automatic
 COMMENT     NCBI gi: 494978
 FEATURES             Location/Qualifiers
      source          1..1080
                      /clone_lib="expression, cDNA"
                      /organism="Homo sapiens"
                      /cell_line="clone 22 (T cell)"
      5'UTR           1..83
      CDS             84..791
```

```
                     /standard_name="FMS-like tyrosine kinase-3 ligand"
                     /note="ligand for the flt3/flk-2 tyrosine kinase receptor;
                      NCBI gi: 494979"
                     /codon_start=1
                     /function="stimulates proliferation of early hematopoietic
                     cells"
                     /product="flt3 ligand"
                     /translation="MTVLAPAWSPTTYLLLLLLLSSGLSGTQDCSFQHSPISSDFAVK
                     IRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTVAGSKMQGLLERVN
                     TEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQP
                     DSSTLPPPWSPRPLEATAPTAPQPPLLLLLLLPVGLLLLAAAWCLHWQRTRRRTPRPG
                     EQVPPVPSPQDLLLVEH"
     sig_peptide     84..161
     misc_feature    162..629
                     /note="extracellular domain"
     misc_feature    630..698
                     /note="transmembrane domain"
     misc_feature    699..788
                     /note="cytoplasmic domain"
     3'UTR           792..1080
     misc_feature    1015..1080
                     /note="ATTTA mRNA instability motif"
     polyA_signal    1059..1064
     polyA_site      1080
                     /note="32 A residues"
BASE COUNT      204 a    384 c    290 g    202 t
ORIGIN

Initial Score    =    378  Optimized Score  =    608  Significance = 26.18
Residue Identity =    70%  Matches          =    644  Mismatches   =   212
Gaps             =     60  Conservative Substitutions               =     0

             X          10          20          30          40          50
             GTC---GACTGGAAC----GAGACGACC-TG-CTCTGTCAC---AGGCATGAGGGGTCCCCG
                ||| |||| |  |    |   |  ||  | | | |  ||  ||||||||||| |||||
   GGGCATGAGGGTCCGAGACTTGTTCTTCTGTCCCTTCCAAGACCCGGCGACAGGAGGCATGAGGGGCCCCCG
      10     X  20          30          40          50          60          70

            60          70          80          90         100         110         120
   GCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGT
   || || ||||||||||||||||||||||||||||||||   |||  |   | || ||| ||||||||||||||
   GCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCAACAACCTATC---TCCTCCTGCTGCTGCTGCTGAGC
    80          90         100         110         120          130         140

        130         140         150         160         170         180         190
   CCTTGCCTGCGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAGTTT
    |  | ||  | ||||| |  ||||| | ||||   ||||| ||||||||||| |||||   || ||  |
   TCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAAAATC
    150         160         170         180         190         200         210

      200         210         220         230         240         250         260
   AGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAGAAGCAC
    | ||| || ||||| |||||||| |||||||||||||||| ||||||  ||| || |||||||| ||| |
   CGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTC
  220         230         240         250         260         270         280         290

    270         280         290         300         310         320         330
   TGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAG
   |||  || | | ||| | ||  |||| || |||||||||| ||||  || |||||||| || ||||| |||
   TGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAG
```



```
          300         310         320         330         340         350         360

    340         350         360         370         380         390         400
    ATGCAAACGCTT-CTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCCTACC
    |||| || |||| ||||||  ||| ||||| ||||||||| |||||||||  |||| |||| |||||||  ||
    ATGC-AAGGCTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCC
          370         380         390         400         410         420         430

  410         420         430         440         450         460         470         480
    AGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCTGCTTGCTCT
       ||||| || ||||||||||||||||||||||| ||||||| |||| |||| |  |||||| | || ||
    CAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCT
        440         450         460         470         480         490         500

          490         500         510         520         530         540         550
    GAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTCCTC
    ||||||||| |||    |  |  |||||| ||||| |||||||||||| ||||||| ||||| |||||||||
    GAAGCCCTGGATC-----ACTC--GCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTC
      510             520         530         540         550         560         570

        560         570         580         590         600         610         620
    CACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCAGCT
     |||||||  |||||| ||||||||     |||| || |||||  ||| || | |||| ||    ||| |  ||
    AACCCTGCCACCCCCATGGAGTCCCCGGCCCCTGGAGGCCAC--AGCCCCGACAGCCCCG---CAGCCCCCT
          580         590         600         610         620           630

      630         640         650         660         670         680         690
    GTTGCTCCTGCTGCTGCTGCTGCCTCTCACACTGGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTGGCAAAG
     |||||||  || ||||||||||  |     ||  ||||||||||  || ||||||  |||| | ||||||  ||
    -CTGCTCCTCCTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCTGGTGCCTGCACTGGCAGAG
    640         650         660         670         680         690         700

    700         710         720         730           740         750         760
    GGCAAGAAGGAGGGGGGAGCTCCACCCTGGGGTGC-----CCCTCCCCTCCC--ATCCCTAGGATTCGAGCCT
    | |  |  |||  ||  | | || |||||||| ||     || ||| ||||   |||| ||||   |  ||
    GACGCGGCGGA--GGACACC-CCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCT
  710         720         730         740         750         760         770

        770         780         790           800           810         820
    TGTGCATCGTTGACTCAGCCAGGGTCTTATC--TCGG----TTACAC----CTGTAATCTCAGCCCT-TGGG
    |||| | |  ||||   |||| || || |||   |||     ||| ||    | || |   ||| | | |
    TGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCTTAAACAACGCAGTGA-GACAGACATCTATC
  780         790         800         810         820         830         840

              830         840         850         860         870         X
    AGCC-------CAGAGCAGGATTGCTGAATGGTCTGGAGCAGGTCGTCTCGTTCCAGTCGAC
    | ||        ||| | |||| | ||||  ||     ||  ||  ||| |   ||||  ||
    ATCCCATTTTACAGGGGAGGA-TACTGA--GGCACACAGAGGGGAGTCACCAGCCAGAGGATGTATAGCCTG
  850         860         870           880         890         900         910
```

5. US-08-162-407-1 (1-879)
   HSC1RCP Human complement C1r mRNA, complete cds.

```
ID   HSC1RCP    standard; RNA; PRI; 2493 BP.
XX
AC   M14058;
XX
DT   02-APR-1988 (Rel. 15, Created)
DT   16-DEC-1994 (Rel. 42, Last updated, Version 3)
```

```
XX
DE   Human complement C1r mRNA, complete cds.
XX
KW   .
XX
OS   Homo sapiens (human)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
XX
RN   [1]
RP   1-2493
RA   Leytus S.P., Kurachi K., Sakariassen K.S., Davie E.W.;
RT   "Nucleotide sequence of the cDNA coding for human complement C1r";
RL   Biochemistry 25:4855-4863(1986).
XX
DR   SWISS-PROT; P00736; C1R_HUMAN.
XX
CC   NCBI gi: 179643
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2493
FT                   /organism="Homo sapiens"
FT   mat_peptide     64..1452
FT                   /note="C1r A chain"
FT                   /codon_start=1
FT   CDS             64..2181
FT                   /gene="C1R"
FT                   /map="12p13"
FT                   /note="human complement C1r; NCBI gi: 179644"
FT                   /codon_start=1
FT                   /gdb_xref="GDB-119-729"
FT   mat_peptide     1453..2178
FT                   /note="C1r B chain serine protease"
FT                   /codon_start=1
XX
SQ   Sequence 2493 BP; 619 A; 680 C; 662 G; 532 T; 0 other;

Initial Score    =    173  Optimized Score  =    397  Significance = 10.71
Residue Identity =    51%  Matches          =    486  Mismatches   =   342
Gaps             =    123  Conservative Substitutions              =     0

            X         10        20        30        40        50
            GTCGACT--GGAACGAGACGACC-TGCTCTG----TCA--CAGGCATG-AGGGGTCCCCGGC
            | | |||   |  |  | || ||  ||| ||    |||   || || | |    | || |||
    CTACGTTGGAGGCTACTTCTGTTCCTGCCGTCCAGGCTATGAGCTTCAGGAAGACAGGCATTCCTGCCAGGC
       570   X   580       590       600       610       620       630

         60          70        80        90       100       110
    AGAGATGA-CA--GTGCTG-GCGC-CAGCCTGGAGCCCAAAT-TCCTCCCTGTTGCTGCTGTTGCTGCTGCT
      |||   | ||  | ||||  | |  || |   || | | || |||  ||||  | | |  | | | || |
    TGAGTGCAGCAGCGAGCTGTACACGGAGGCATCAGGCTACATCTCCAGCCTGGAG-TACCCTCGGTCCTAC-
      640       650       660       670       680       690       700

   120       130          140       150       160       170       180
    GAGTCCTTGCCTGCGGGG-----GACACCTGACTGTT--ACTTCAGCCACAGTCCCATCTCCTCCAACTTCAA
        |||  ||||||  |      ||| |   | | |  |      ||| ||  |||  | ||| ||| |||
    -CCCCCTGACCTGCGCTGCAACTACAGCATCCGGGTGGAGCGGGGCCTCA--CCCTGCACCT-CAAGTTC--
       710       720       730       740       750       760       770
```



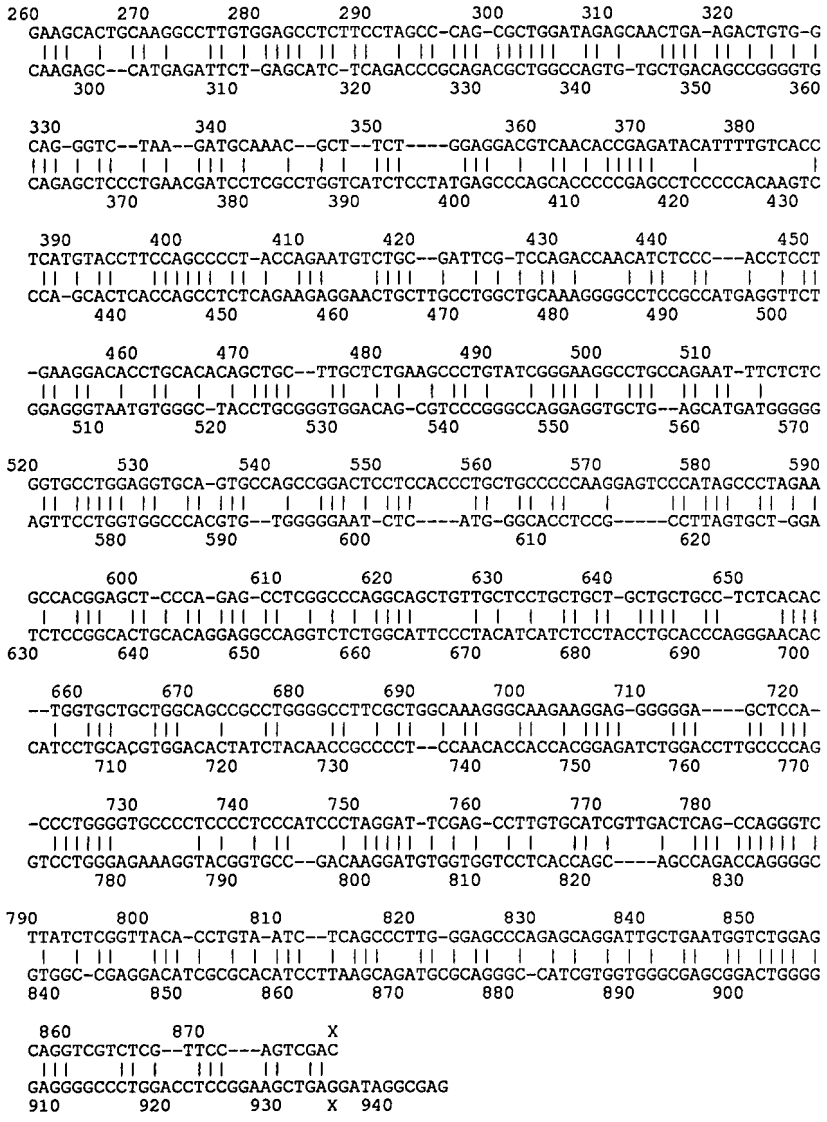
```
     190       200       210       220       230       240       250
AGTGAAG--TTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTC----AATCT
  || ||  ||| || | |||| |||||||| ||   |||  ||| | | | || || ||  |    ||||
-CTGGAGCCTTTTGATA-TTGA-TGACCACCAGC---AAG--TACACTGCCCCTATGACCAGCTACAGATCT
      780       790       800       810       820       830

        260       270       280       290       300       310
-TCAGGACGAGAAG--CACTG-CAAGGCCT-TGTGGAGCCTCTTCCTAGCCCAGCGCTGGATA-GAGCAACT
 |    ||| ||||  || || | ||  || || | |||     ||   ||| |  || || |  |||| |
ATGCCAACGGGAAGAACATTGGCGAGTTCTGTGGGAAGCAAAGGCC---CCCCGACCTCGACACCAGCAGC-
  840       850       860       870       880       890       900

 320       330       340       350       360       370       380
GAAGACTGTGG-CAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCT
 ||  ||||||    | | |       || | |  || || ||||  |      |  ||  |       ||||
-AATGCTGTGGATCTGCTGTTCTTCACAGATGAGTC-GGGGGACAGCCGGGGCTGGAAGCTGCGCTACACCA
     910       920       930       940       950       960       970

390       400       410       420       430       440       450
CATGTACCTTCCA--GCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGA
|    | | || |  ||||| | |  || | |  |  | | ||  || ||||  |||   |||| ||
CCGAGATCATCAAGTGCCCCCAGCCCAA-GAC--C--CTAGACGAGTTCACCATCATCCAGAACCTGCAG--
    980       990      1000      1010      1020      1030

460       470       480       490       500       510       520
CACCTGCACACAGCTGCTTG-CT-CTGAA--GC--CCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTG
|  | | || ||| | | || || ||  |  ||  |||| |   | |||| || ||||    |    ||
CCTCAGTAC-CAGTTCCGTGACTACTTCATTGCTACCTGCA--AGCAAGG-CTACCAGCTCATAGAGGGGAA
1040      1050      1060      1070      1080      1090      1100

     530       540       550       560       570       580       590
CCTGGAGGTGCAGTGC--CAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGA--GTCCCATAGCCCTAGAAG
|| || | |||| | |  || | |   ||  |||   || || | |   | |  || |||| |||| ||| |
CCAGGTGCTGCATTCCTTCA-CAGCTGTCTGCCAGGATGATGGCACGTGGCATCGTGCCAT-GCCC-AGATG
1110      1120      1130      1140      1150      1160      1170

      600       610       620       630       640       650
CCACGGAGCTCCCAGAGCCT---CGGCCCAGGCAGCTGTTGCTCCTGCTG-CTGCTGCTGC-CTCTCACACT
| |  ||  ||   ||   |  | |||| |  | |||   ||  || || || || | | | | |   |||| |
CAA--GA--TCAAGGACTGTGGGCAGCCCCGAAACCTG--CCTAATGGTGACTTCCGTTACACCACCACAAT
   1180      1190      1200      1210      1220      1230      1240

  660       670       680       690       700       710
G---GTG--CTGCTGGCAG--CCGCCTGGGG-CCTTC-GC--TG-GC-AAAGGGCAAGAAGGAGGGGGGAGC
|   |||  |  ||   ||  |||  |   | || | ||  || || | |  ||||| | ||   ||||
GGGAGTGAACACCTACAAGGCCCGTATCCAGTACTACTGCCATGAGCCATATTACAAGATGCAGACCAGAGC
     1250      1260      1270      1280      1290      1300      1310

 720       730       740       750       760       770       780
T-CCACCCTGGGG--TGCCCCTCCCCT---CCCATCCCTAGGATTCGAGCCTTGTGCATCGTTGACTCAGCC
|  || |  || |  ||  |       |   | | | |  || |   | || || || |    |||  |||
TGGCAGCAGGGAGTCTGAGCAAGGGGTGTACACCTGCACAGCA-CAGGGCATT-TGGAAGAATGA-ACAG-A
   1320      1330      1340      1350      1360      1370      1380

     790       800       810       820       830       840       850
AGGGTCTTATCTCGGTTACACCTGTAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGA
||||    |   | || | |       |  |||    |||||  || | | |      |  |||  |   | |
AGGGAGAGA---AGATTCCTCGGTGCTTGCCAGTGTGTGGGAAGCCCGTG-AACCCCGTGGAACAG-AGGCA
     1390      1400      1410      1420      1430      1440
```

```
      860        870        X
    GCAGGTCGTC-TCGTTCCAGTCGAC
    ||   |  ||   |  | |   | |
    GCGCATAATCGGAGGGCAAAAAGCCAAGATGGGCA
  1450      1460       1470  X   1480
```

6. US-08-162-407-1 (1-879)
   HUMIRBPG Human interstitial retinol-binding protein (IRBP)

```
LOCUS       HUMIRBPG     9711 bp ds-DNA             PRI       06-JAN-1995
DEFINITION  Human interstitial retinol-binding protein (IRBP) gene, complete
            cds.
ACCESSION   J05253
KEYWORDS    interstitial retinol-binding protein.
SOURCE      Human DNA, clone HGL.3.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 9711)
  AUTHORS   Fong,S.L., Fong,W.B., Morris,T.A., Kedzie,K.M. and Bridges,C.D.
  TITLE     Characterization and comparative structural features of the gene
            for human interstitial retinol-binding protein
  JOURNAL   J. Biol. Chem. 265, 3648-3653 (1990)
  STANDARD  full automatic
COMMENT     NCBI gi: 186534
FEATURES             Location/Qualifiers
     source          1..9711
                     /organism="Homo sapiens"
                     /cell_type="leukocyte"
                     /sequenced_mol="DNA"
                     /clone="HGL3"
     prim_transcript 1..>9505
                     /note="IRBP mRNA and introns (alt.)"
     prim_transcript 7..>9505
                     /note="IRBP mRNA and introns"
     prim_transcript 15..>9505
                     /note="IRBP mRNA and introns (alt.)"
     CDS             join(130..3180,4966..5156,7017..7159,8766..9121)
                     /gene="IRBP"
                     /map="chromosome 10"
                     /note="precursor;  NCBI gi: 386835"
                     /codon_start=1
                     /product="interstitial retinol-binding protein"
                     /translation="MREWVLLMSVLLCGLAGPTHLFQPSLVLDMAKVLLDNYCFPENL
                     LGMQEAIQQAIKSHEILSISDPQTLASVLTAGVQSSLNDPRLVISYEPSTPEPPPQVP
                     ALTSLSEEELLAWLQRGLRHEVLEGNVGYLRVDSVPGQEVLSMMGEFLVAHVWGNLMG
                     TSALVLDLRHCTGGQVSGIPYIISYLHPGNTILHVDTIYNRPSNTTTEIWTLPQVLGE
                     RYGADKDVVVLTSSQTRGVAEDIAHILKQMRRAIVVGERTGGGALDLRKLRIGESDFF
                     FTVPVSRSLGPLGGGSQTWEGSGVLPCVGTPAEQALEKALAILTLRSALPGVVHCLQE
                     VLKDYYTLVDRVPTLLQHLASMDFSTVVSEEDLVTKLNAGLQAASEDPRLLVRAIGPT
                     ETPSWPAPDAAAEDSPGVAPELPEDEAIRQALVDSVFQVSVLPGNVGYLRFDSFADAS
                     VLGVLAPYVLRQVWEPLQDTEHLIMDLRHNPGGPSSAVPLLLSYFQGPEAGPVHLFTT
                     YDRRTNITQEHFSHMELPGPRYSTQRGVYLLTSHRTATAAEEFAFLMQSLGWATLVGE
                     ITAGNLLHTRTVPLLDTPEGSLALTVPVLTFIDNHGEAWLGGGVVPDAIVLAEEALDK
                     AQEVLEFHQSLGALVEGTGHLLEAHYARPEVVGQTSALLRAKLAQGAYRTAVDLESLA
                     SQLTADLQEVSGDHRLLVFHSPGELVVEEAPPPPPAVPSPEELTYLIEALFKTEVLPG
                     QLGYLRFDAMAELETVKAVGPQLVRLVWQQLVDTAALVIDLRYNPGSYSTAIPLLCSY
                     FFEAEPRQHLYSVFDRATSKVTEVWTLPQVAGQRYGSHKDLYILMSHTSGSAAEAFAH
                     TMQDLQRATVIGEPTAGGALSVGIYQVGSSPLYASMPTQMAMSATTGKAWDLAGVEPD
                     ITVPMSEALSIAQDIVALRAKVPTVLQTAGKLVADNYASAELGAKMATKLSGLQSRYS
                     RVTSEVALAEILGADLQMLSGDPHLKAAHIPENAKDRIPGIVPMQIPSPEVFEELIKF
                     SFHTNVLEDNIGYLRFDMFGDGELLTQVSRLLVEHIWKKIMHTDAMIIDMRFNIGGPT
                     SSIPILCSYFFDEGPPVLLDKIYSRPDDSVSELWTHAQVVGERYGSKKSMVILTSSVT
                     AGTAEEFTYIMKRLGRALVIGEVTSGGCQPPQTYHVDDTNLYLTIPTARSVGASDGSS
                     WEGVGVTPHVVVPAEEALARAKEMLQHNQLRVKRSPGLQDHL"
     exon            <130..3180
                     /gene="IRBP"
                     /number=1
     sig_peptide     130..177
                     /gene="IRBP"
                     /codon_start=1
     mat_peptide     join(178..3180,4966..5156,7017..7159,8766..9118)
                     /gene="IRBP"
                     /codon_start=1
                     /product="interstitial retinol-binding protein"
     intron          3181..4965
                     /note="IRBP intron A"
     exon            4966..5156
                     /gene="IRBP"
                     /number=2
     intron          5157..7016
                     /note="IRBP intron B"
     exon            7017..7159
                     /number=3
     intron          7160..8765
                     /note="IRBP intron C"
     exon            8766..>9121
                     /gene="IRBP"
                     /number=4
BASE COUNT     2246 a   2672 c   2641 g   2152 t
ORIGIN      Chromosome 10q11.2.
```

```
Initial Score    =    172  Optimized Score  =    391  Significance = 10.64
Residue Identity =    50%  Matches          =    481  Mismatches   =   348
Gaps             =    125  Conservative Substitutions               =     0

            X       10        20        30        40        50
            GTCGACTGGAACGAGACGACCTGCTCTG-TCACAGGCATGAGGGGTCCCCGG--CAGAGA-T
            |   |  ||    || | | |||| | |  ||  | || |    |  | | | |  ||| ||
CAGCTGAGAAGGACAAGGGCGGAAGGC-AGCTGCACAGAGCAGGGCCACGGCTTGCACACAGTCCAGGGAGC
       30 X      40        50        60        70        80        90

60        70        80        90        100       110       120
GACAGTGCTGGCGCCA-GCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGA-GTCCTTGC
    |||| || |||| ||||    |||    ||| || || ||       | | | |||| | |||| |||
TTTTGTGCAGGAGCCAGGCCT---CCCCCTGGTCC-CCATGATGAGAGAATGGGTTCTGCTCATGTCCGTGC
     100       110       120       130       140       150       160

   130           140       150       160       170       180       190
-----CTGCGG---GG--GACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAG
    ||| ||   ||  | | ||  ||   |  | |||||| ||  ||   | ||  |    ||  |  |||
TGCTCTGTGGCCTGGCTGGCCCCACACACCTGTTCCAGCCA-AG--CCTGGTGCTGGA----CATGGCCAAG
       170       180       190       200       210           220

      200       210       220       230       240       250
TTTAGAGAGTTGACTGACCACCTGCTT-AAAGATTACC--C-AGTCACTGTGGCCGTCAATCTTCAGGACGA
  |       |||  | || | ||||||    ||  |||  |  | ||    ||  | ||  |  |||| |
GT---CCTCTTGGATAACTA-CTGCTTCCCGGAGAACCTGCTGGGCATGCAGGAAGCCATCCAGCAGGCCAT
     230       240       250       260       270       280       290
```



7. US-08-162-407-1 (1-879)
   HUMIRBPG Human interstitial retinol-binding protein (IRBP)

```
LOCUS       HUMIRBPG     9711 bp ds-DNA             PRI       14-SEP-1994
DEFINITION  Human interstitial retinol-binding protein (IRBP) gene, complete
            cds.
ACCESSION   J05253
KEYWORDS    interstitial retinol-binding protein.
SOURCE      Human DNA, clone HGL.3.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 9711)
  AUTHORS   Fong,S.-L., Fong,W.-B., Morris,T.A., Kedzie,K.M. and Bridges,C.D.B.
  TITLE     Characterization and comparative structural features of the gene
            for human interstitial retinol-binding protein
  JOURNAL   J. Biol. Chem. 265, 3648-3653 (1990)
  STANDARD  full automatic
COMMENT     NCBI gi: 186534
FEATURES             Location/Qualifiers
     source          1..9711
                     /organism="Homo sapiens"
                     /cell_type="leukocyte"
                     /sequenced_mol="DNA"
                     /clone="HGL3"
     prim_transcript 1..>9505
                     /note="IRBP mRNA and introns (alt.)"
     prim_transcript 7..>9505
                     /note="IRBP mRNA and introns"
     prim_transcript 15..>9505
                     /note="IRBP mRNA and introns (alt.)"
     CDS             join(130..3180,4966..5156,7017..7159,8766..9121)
                     /gene="IRBP"
                     /map="chromosome 10"
                     /note="precursor;  NCBI gi: 386835"
                     /codon_start=1
                     /product="interstitial retinol-binding protein"
                     /translation="MREWVLLMSVLLCGLAGPTHLFQPSLVLDMAKVLLDNYCFPENL
                     LGMQEAIQQAIKSHEILSISDPQTLASVLTAGVQSSLNDPRLVISYEPSTPEPPPQVP
                     ALTSLSEEELLAWLQRGLRHEVLEGNVGYLRVDSVPGQEVLSMMGEFLVAHVWGNLMG
                     TSALVLDLRHCTGGQVSGIPYIISYLHPGNTILHVDTIYNRPSNTTTEIWTLPQVLGE
                     RYGADKDVVVLTSSQTRGVAEDIAHILKQMRRAIVVGERTGGGALDLRKLRIGESDFF
                     FTVPVSRSLGPLGGGSQTWEGSGVLPCVGTPAEQALEKALAILTLRSALPGVVHCLQE
                     VLKDYYTLVDRVPTLLQHLASMDFSTVVSEEDLVTKLNAGLQAASEDPRLLVRAIGPT
                     ETPSWPAPDAAAEDSPGVAPELPEDEAIRQALVDSVFQVSVLPGNVGYLRFDSFADAS
                     VLGVLAPYVLRQVWEPLQDTEHLIMDLRHNPGGPSSAVPLLLSYFQGPEAGPVHLFTT
                     YDRRTNITQEHFSHMELPGPRYSTQRGVYLLTSHRTATAAEEFAFLMQSLGWATLVGE
                     ITAGNLLHTRTVPLLDTPEGSLALTVPVLTFIDNHGEAWLGGGVVPDAIVLAEEALDK
                     AQEVLEFHQSLGALVEGTGHLLEAHYARPEVVGQTSALLRAKLAQGAYRTAVDLESLA
                     SQLTADLQEVSGDHRLLVFHSPGELVVEEAPPPPPAVPSPEELTYLIEALFKTEVLPG
                     QLGYLRFDAMAELETVKAVGPQLVRLVWQQLVDTAALVIDLRYNPGSYSTAIPLLCSY
                     FFEAEPRQHLYSVFDRATSKVTEVWTLPQVAGQRYGSHKDLYILMSHTSGSAAEAFAH
                     TMQDLQRATVIGEPTAGGALSVGIYQVGSSPLYASMPTQMAMSATTGKAWDLAGVEPD
                     ITVPMSEALSIAQDIVALRAKVPTVLQTAGKLVADNYASAELGAKMATKLSGLQSRYS
                     RVTSEVALAEILGADLQMLSGDPHLKAAHIPENAKDRIPGIVPMQIPSPEVFEELIKF
                     SFHTNVLEDNIGYLRFDMFGDGELLTQVSRLLVEHIWKKIMHTDAMIIDMRFNIGGPT
                     SSIPILCSYFFDEGPPVLLDKIYSRPDDSVSELWTHAQVVGERYGSKKSMVILTSSVT
                     AGTAEEFTYIMKRLGRALVIGEVTSGGCQPPQTYHVDDTNLYLTIPTARSVGASDGSS
```

```
                     WEGVGVTPHVVVPAEEALARAKEMLQHNQLRVKRSPGLQDHL"
     exon            <130..3180
                     /gene="IRBP"
                     /number=1
     sig_peptide     130..177
                     /gene="IRBP"
                     /codon_start=1
     mat_peptide     join(178..3180,4966..5156,7017..7159,8766..9118)
                     /gene="IRBP"
                     /codon_start=1
                     /product="interstitial retinol-binding protein"
     intron          3181..4965
                     /note="IRBP intron A"
     exon            4966..5156
                     /gene="IRBP"
                     /number=2
     intron          5157..7016
                     /note="IRBP intron B"
     exon            7017..7159
                     /number=3
     intron          7160..8765
                     /note="IRBP intron C"
     exon            8766..>9121
                     /gene="IRBP"
                     /number=4
BASE COUNT     2246 a   2672 c   2641 g   2152 t
ORIGIN      Chromosome 10q11.2.

Initial Score    =    172  Optimized Score  =    391  Significance = 10.64
Residue Identity =    50%  Matches          =    481  Mismatches   =   348
Gaps             =    125  Conservative Substitutions               =     0

             X       10        20        30        40        50
             GTCGACTGGAACGAGACGACCTGCTCTG-TCACAGGCATGAGGGGTCCCCGG--CAGAGA-T
    CAGCTGAGAAGGACAAGGGCGGAAGGC-AGCTGCACAGAGCAGGGCCACGGCTTGCACACAGTCCAGGGAGC
          30 X      40        50        60        70        80        90

    60        70        80        90       100       110       120
    GACAGTGCTGGCGCCA-GCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGA-GTCCTTGC
    TTTTGTGCAGGAGCCAGGCCT---CCCCCTGGTCC-CCATGATGAGAGAATGGGTTCTGCTCATGTCCGTGC
        100       110       120       130       140       150       160

       130       140       150       160       170       180       190
    ----CTGCGG---GG--GACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAG
    TGCTCTGTGGCCTGGCTGGCCCCACACACCTGTTCCAGCCA-AG--CCTGGTGCTGGA----CATGGCCAAG
           170       180       190       200       210       220

         200       210       220       230       240       250
    TTTAGAGAGTTGACTGACCACCTGCTT-AAAGATTACC--C-AGTCACTGTGGCCGTCAATCTTCAGGACGA
    GT---CCTCTTGGATAACTA-CTGCTTCCCGGAGAACCTGCTGGGCATGCAGGAAGCCATCCAGCAGGCCAT
        230       240       250       260       270       280       290

  260       270       280       290       300       310       320
    GAAGCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCC-CAG-CGCTGGATAGAGCAACTGA-AGACTGTG-G
    CAAGAGC--CATGAGATTCT-GAGCATC-TCAGACCCGCAGACGCTGGCCAGTG-TGCTGACAGCCGGGGTG
```



```
      300       310       320       330       340       350       360

  330       340       350       360       370       380
  CAG-GGTC--TAA--GATGCAAAC--GCT--TCT----GGAGGACGTCAACACCGAGATACATTTTGTCACC
  CAGAGCTCCCTGAACGATCCTCGCCTGGTCATCTCCTATGAGCCCAGCACCCCCGAGCCTCCCCCACAAGTC
        370       380       390       400       410       420       430

   390       400       410       420       430       440       450
  TCATGTACCTTCCAGCCCCT-ACCAGAATGTCTGC--GATTCG-TCCAGACCAACATCTCCC---ACCTCCT
  CCA-GCACTCACCAGCCTCTCAGAAGAGGAACTGCTTGCCTGGCTGCAAAGGGGCCTCCGCCATGAGGTTCT
       440       450       460       470       480       490       500

       460       470       480       490       500       510
  -GAAGGACACCTGCACACAGCTGC--TTGCTCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAAT-TTCTCTC
  GGAGGGTAATGTGGGC-TACCTGCGGGTGGACAG-CGTCCCGGGCCAGGAGGTGCTG--AGCATGATGGGGG
      510       520       530       540       550       560       570

520       530       540       550       560       570       580       590
  GGTGCCTGGAGGTGCA-GTGCCAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAA
  AGTTCCTGGTGGCCCACGTG--TGGGGGAAT-CTC----ATG-GGCACCTCCG-----CCTTAGTGCT-GGA
       580       590       600       610       620

       600       610       620       630       640       650
  GCCACGGAGCT-CCCA-GAG-CCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCT-GCTGCTGCC-TCTCACAC
  TCTCCGGCACTGCACAGGAGGCCAGGTCTCTGGCATTCCCTACATCATCTCCTACCTGCACCCAGGGAACAC
630       640       650       660       670       680       690       700

   660       670       680       690       700       710       720
  --TGGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGGAG-GGGGGA----GCTCCA-
  CATCCTGCACGTGGACACTATCTACAACCGCCCCT--CCAACACCACCACGGAGATCTGGACCTTGCCCCAG
       710       720       730       740       750       760       770

       730       740       750       760       770       780
  -CCCTGGGGTGCCCCTCCCCTCCCATCCCTAGGAT-TCGAG-CCTTGTGCATCGTTGACTCAG-CCAGGGTC
  GTCCTGGGAGAAAGGTACGGTGCC--GACAAGGATGTGGTGGTCCTCACCAGC----AGCCAGACCAGGGGC
       780       790       800       810       820       830

790       800       810       820       830       840       850
  TTATCTCGGTTACA-CCTGTA-ATC--TCAGCCCTTG-GGAGCCCAGAGCAGGATTGCTGAATGGTCTGGAG
  GTGGC-CGAGGACATCGCGCACATCCTTAAGCAGATGCGCAGGGC-CATCGTGGTGGGCGAGCGGACTGGGG
  840       850       860       870       880       890       900

   860       870       X
  CAGGTCGTCTCG--TTCC---AGTCGAC
  GAGGGGCCCTGGACCTCCGGAAGCTGAGGATAGGCGAG
  910       920       930    X  940
```

8. US-08-162-407-1 (1-879)
   HSCGPDEG2 Human rod cGMP-phosphodiesterase gamma-subunit (PD

```
LOCUS       HSCGPDEG2    2417 bp    DNA             PRI       20-DEC-1993
DEFINITION  Human rod cGMP-phosphodiesterase gamma-subunit (PDEG) gene, exon 2.
ACCESSION   U00481
KEYWORDS    .
SEGMENT     2 of 3
SOURCE      human.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2417)
  AUTHORS   Hahn,L.B., Berson,E.L. and Dryja,T.P.
  TITLE     Evaluation of the gene encoding the gamma subunit of rod
            phosphodiesterase in retinitis pigmentosa
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 2417)
  AUTHORS   Cotran,P.R., Bruns,G.A., Berson,E.L. and Dryja,T.P.
  TITLE     Genetic analysis of patients with retinitis pigmentosa using a
            cloned cDNA probe for the human gamma subunit of cyclic GMP
            phosphodiesterase
  JOURNAL   Exp. Eye Res. 53, 557-564 (1991)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 2417)
  AUTHORS   Tuteja,N., Danciger,M., Klisak,I., Tuteja,R., Inana,G.,
            Mohandas,T., Sparkes,R.S. and Farber,D.B.
  TITLE     Isolation and characterization of cDNA encoding the gamma-subunit
            of cGMP phosphodiesterase in human retina
  JOURNAL   Gene 88, 227-232 (1990)
  STANDARD  full automatic
REFERENCE   4  (bases 1 to 2417)
  AUTHORS   Dryja,T.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1993) Thaddeus P. Dryja, Massachusetts Eye and
            Ear Infirmary, Harvard Medical School, Room CB621, 243 Charles
            Street, Boston, MA 02114-3096, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 409400
FEATURES             Location/Qualifiers
     source          1..2417
                     /map="17q(approx.21.1)"
                     /organism="Homo sapiens"
                     /cell_line="normal human lymphoblast line"
     repeat_unit     1..236
                     /rpt_family="Alu-class IV(sense)"
                     /evidence=experimental
     repeat_unit     321..601
                     /rpt_family="Alu-class II(sense)"
                     /evidence=experimental
     repeat_unit     complement(886..1173)
                     /rpt_family="Alu-classIII(sense)"
                     /evidence=experimental
     exon            1637..1841
                     /gene="PDEG"
                     /note="1637..1695 is noncoding; initiation codon at 1696"
                     /number=2
                     /evidence=experimental
                     /product="gamma-subunit of rod cGMP phosphodiesterase"
     misc_feature    1682..1683
                     /note="C deletion polymorphism"
     repeat_unit     2088..2371
```



```
                     /rpt_family="Alu-classII(sense)"
                     /evidence=experimental
BASE COUNT      621 a    598 c    810 g    388 t
ORIGIN

Initial Score    =    162  Optimized Score  =    392  Significance =  9.88
Residue Identity =    49%  Matches          =    462  Mismatches   =   376
Gaps             =     97  Conservative Substitutions               =     0

             X       10        20        30        40        50
             GTCGACTGGAACGAGACGACC--TGCTCTGTCACAGGCATGAGGGGTC-CCCGGCAGAGATG
              ||  ||||  |  |  | ||  | | | | ||   ||   || || |   | |||||| |
   AGGCAGGCTCCTCTGCTGGGCCTGGCAGGCCCATTCCCAG-CAGGAGCCACAGAGGACTTTCAGCAGAG-TA
 810       820       830       840       850       860       870

  60          70        80        90        100       110
   ACAG-TG---CTGGCGC-CAGCCTGGAGCC-CAAATTCC-TCCCTGT-TGCTGCTGTTGCTGCTG----CTG
    ||| ||   | || ||   | ||   |||   ||| ||  | || |  |  ||||  || | ||     ||
   GCAGCTGGGCCGGGTGCGGTGGCTCACGCCTGTAATACCAGCACTTTGGGAGGCTGAGGCGGGTGGATCATG
  880       890       900       910       920       930       940       950

 120       130       140       150       160       170       180       190
   AGTCCTTGCCTGCGGGGGACACCT-GACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAA
   ||  |  |  | || |  | ||||  | |       | |  | | |||   ||| ||  || | ||||   ||
   AGGTCAGGAGTTCGAGACA-ACCTCGTCAACATGGTGAAACCCCGTC---TCTACTAAAAATGCAAA---AA
          960       970       980       990       1000      1010

          200       210       220       230       240       250       260
   GTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTAC-CCAGTCACTGTGGCCGTCAATCTTCAGG-ACGAG
    |  | |||·| | | |    | || |   |||   |   | || | |  || |   |   | ||| | |||
   -TGCAAAGATTAGCCGG---GCATGGTGGCAGACGCCGGTAATCCCAGCTGCTGGGGAGGCTGAGGCAGGAG
     1020      1030      1040      1050      1060      1070      1080

          270       280       290       300       310       320       330
   AAGCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGG
   || ||||  ||  ||  | || |    || | ||   ||| |  ||||| || |||| |  |     ||||
   AATCACTTGAA-CCCGGGAGGCGGAGGTTGC-AG-TGAGC-CGAGATAGTGCCACTGCACTCCAGCCCAGGC
      1090      1100      1110      1120      1130      1140      1150

        340       350       360       370       380       390       400
   TCTAAGATGCAAACGCT-TCTGGAGGAC--GTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAG
      ||||   |||| || |||  |  ||    ||| ||  | | | |    |       ||  ||  | ||
   AACAAGAGCAAAACTCTGTCTCAAAAACAAAACAAAACAAAAAAAAAACCCAG-AGTAGCAGTTAAATGCCGA
        1160      1170      1180      1190      1200      1210      1220

         410       420       430       440       450       460
   CCCCTACCA-GAATGTCTGCGATTCG---TCCAGACCAACATCTCCCACCTCCTGA-AGGACACCTGCACAC
   |       || |  |  ||| | |  |     |||   ||        |  |  | |  || ||||  ||||
   CGGGAGTCACGCTTCCCTGGGGTGGGGAACGCAGGGGAAGGGGCAGGAAGTTGTCAGTGGTCACC--CACAG
       1230      1240      1250      1260      1270      1280      1290

  470       480       490       500       510       520       530
   AGCTGCTTGCTCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGT-GCCTGGAGGTGCAGTGC
    |  |  || |  | |  || |  || |||   |||||||  |  ||      | | |  | |  ||| | ||
   GGAGG--TGGTGAGGACTCCAGGCTCAGGA---CCTGCCACAGGTTGGGAGGCTAGGTTTGCTGTGGATTGT
         1300      1310      1320      1330      1340      1350

 540       550       560       570       580       590       600
   CAGCC-GGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCT-AGAAGCCA-CGGAGCTCCCAGAG
     ||| || |        || |  | |     ||||| ||  |||  |  |  |||| ||  ||     | |
```

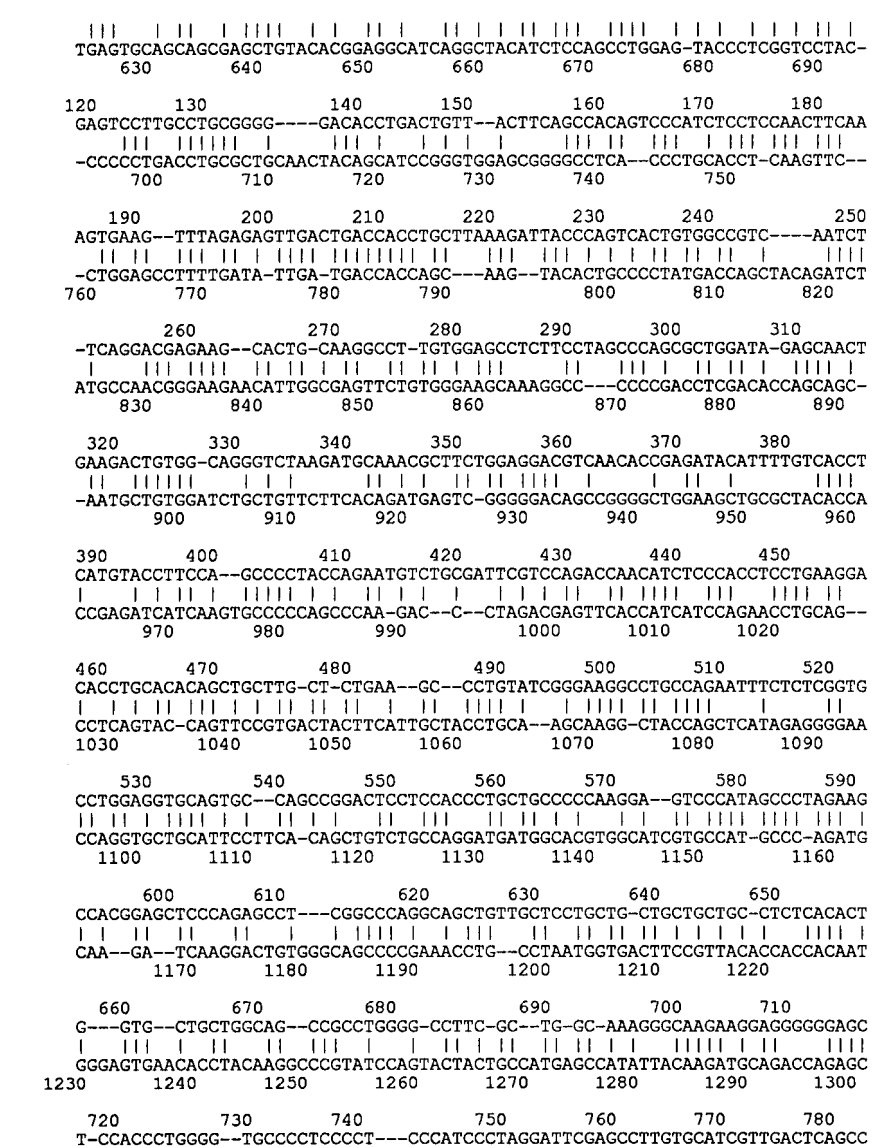
```
    GGGCCGGGGCGAGGGAGGCCAG--GACAGGGGGGAGT-CCGGAGCTGTGTGTGGCCAGCGTTGC-AGGGGGG
1360       1370       1380       1390       1400       1410       1420

  610        620        630        640        650        660        670
   CCTCGGCCCAGGCAGCTGTTGCTCCTGC-TGCTGCTGCTGCCTCTCACA-CTGGTGCTGCTGGCAGCCGCCT
   | ||   | |||||   |         || ||| | ||   ||   || | ||||   |||||   || ||
   CATC-CACTAGGCAAGGGGAATGGGAGCGTGCGGGTGGACCCAAGCAAAGCTGG-CAGGCTGGTTACC-CCA
 1430       1440       1450       1460       1470       1480       1490

  680        690        700        710        720        730        740
   GGGGCCTTCGCTGGCAAAG-GGCAAGAAGGAGGGGGGAGCT--CCACCCTGGGGTGCCCCTCCCCTCCCATC
   |||    | || ||    |  |   | | || |  ||||||  |    |||||.  | |  ||||| ||| |
   GGGCGGATTGCAGGGGCTGCTGTCTGGATGACGCTGGAGCTGGCGGGGGCTGGG--TCTCAGCCCCT-CCACC
   1500       1510       1520       1530       1540       1550       1560

   750        760                   770        780        790        800
   CCTAGGATTCGAGCC---------TTG--TGC---ATCGTTGACTCAG-CCAGGGTCTTATCTCGGTTACAC
     | || |  || ||        |||   ||   |  | |||||  | || ||| ||  ||    |  | |
   GGTGGGTTGAGACCCCCACAAGGGTTGGAAGCAGGAGGGCTGACTGGGACCCGGGCCTGGTC----TGCCCC
    1570       1580       1590       1600       1610       1620       1630

      810        820        830        840        850        860        870
   CTGTAATCTCAGCCCTTGGGAGCCCA-GAG-CAGGATTGCTGAATGGTCTGGAGCAG-GTCGTCTCGTTCC-
   | |  | | ||||    ||||| ||| ||| || | |||| |    ||||    |||  ||  |  |  ||
   CAG--ACCGCAGCAGGAGGGAGTCCAGGAGCCAAGGTTGCCGCGGTGTCTCCGTCAGCCTCACCATGAACCT
        1640       1650       1660       1670       1680       1690       1700

          X
   --AGTCGAC
     |  || |
   GGAACCGCCCAAGGCTGAG
      1710 X    1720
```

9. US-08-162-407-1 (1-879)
HSC1R Human mRNA for complement component C1r.

```
LOCUS       HSC1R        2386 bp    RNA             PRI       12-SEP-1993
DEFINITION  Human mRNA for complement component C1r.
ACCESSION   X04701
KEYWORDS    complement protein C1r.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
            Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 2386)
  AUTHORS   Journet,A. and Tosi,M.
  TITLE     Cloning and sequencing of full-length cDNA encoding the precursor
            of human complement component C1r
  JOURNAL   Biochem. J. 240, 783-787 (1986)
  STANDARD  full automatic
COMMENT     NCBI gi: 29538
FEATURES             Location/Qualifiers
     source          1..2386
                     /organism="Homo sapiens"
     CDS             52..102
                     /note="signal peptide (AA -17 to -1);  NCBI gi: 29540"
                     /codon_start=1
                     /translation="MWLLYLLVPALFCRAGG"
     CDS             52..2169
```



```
                     /note="precursor of C1r (AA -17 to 688);  NCBI gi: 29539"
                     /codon_start=1
                     /translation="MWLLYLLVPALFCRAGGSIPIPQKLFGEVTSPLFPKPYPNNFET
                     TTVITVPTGYRVKLVFQQFDLEPSEGCFYDYVKISADKKSLGRFCGQLGSPLGNPPGK
                     KEFMSQGNKMLLTFHTDFSNEENGTIMFYKGFLAYYQAVDLDECASRSKLGEEDPQPQ
                     CQHLCHNYVGGYFCSCRPGYELQEDRHSCQAECSSELYTEASGYISSLEYPRSYPPDL
                     RCNYSIRVERGLTLHLKFLEPFDIDDHQQVHCPYDQLQIYANGKNIGEFCGKQRPPDL
                     DTSSNAVDLLFFTDESGDSRGWKLRYTTEIIKCPQPKTLDEFTIIQNLQPQYQFRDYF
                     IATCKQGYQLIEGNQVLHSFTAVCQDDGTWHRAMPRCKIKDCGQPRNLPNGDFRYTTT
                     MGVNTYKARIQYYCHEPYYKMQTRAGSRESEQGVYTCTAQGIWKNEQKGEKIPRCLPV
                     CGKPVNPVEQRQRIIGGQKAKMGNFPWQVFTNIHGRGGGALLGDRWILTAAHTLYPKE
                     HEAQSNASLDVFLGHTNVEELMKLGNHPIRRVSVHPDYRQDESYNFEGDIALLELENS
                     VTLGPNLLPICLPDNDTFYDLGLMGYVSGFGVMEEKIAHDLRFVRLPVANPQACENWL
                     RGKNRMDVFSQNMFCAGHPSLKQDACQGDSGGVFAVRDPNTDRWVATGIVSWGIGCSR
                     GYGFYTKVLNYVDWIKKEMEEED"
     CDS             103..2169
                     /note="C1r zymogenic form (AA 1-688);  NCBI gi: 579717"
                     /codon_start=1
                     /translation="XIPIPQKLFGEVTSPLFPKPYPNNFETTTVITVPTGYRVKLVFQ
                     QFDLEPSEGCFYDYVKISADKKSLGRFCGQLGSPLGNPPGKKEFMSQGNKMLLTFHTD
                     FSNEENGTIMFYKGFLAYYQAVDLDECASRSKLGEEDPQPQCQHLCHNYVGGYFCSCR
                     PGYELQEDRHSCQAECSSELYTEASGYISSLEYPRSYPPDLRCNYSIRVERGLTLHLK
                     FLEPFDIDDHQQVHCPYDQLQIYANGKNIGEFCGKQRPPDLDTSSNAVDLLFFTDESG
                     DSRGWKLRYTTEIIKCPQPKTLDEFTIIQNLQPQYQFRDYFIATCKQGYQLIEGNQVL
                     HSFTAVCQDDGTWHRAMPRCKIKDCGQPRNLPNGDFRYTTTMGVNTYKARIQYYCHEP
                     YYKMQTRAGSRESEQGVYTCTAQGIWKNEQKGEKIPRCLPVCGKPVNPVEQRQRIIGG
                     QKAKMGNFPWQVFTNIHGRGGGALLGDRWILTAAHTLYPKEHEAQSNASLDVFLGHTN
                     VEELMKLGNHPIRRVSVHPDYRQDESYNFEGDIALLELENSVTLGPNLLPICLPDNDT
                     FYDLGLMGYVSGFGVMEEKIAHDLRFVRLPVANPQACENWLRGKNRMDVFSQNMFCAG
                     HPSLKQDACQGDSGGVFAVRDPNTDRWVATGIVSWGIGCSRGYGFYTKVLNYVDWIKK
                     EMEEED"
     misc_feature    424..426
                     /note="N-linked glycosylation site"
     misc_feature    550..552
                     /note="N-linked glycosylation site"
     misc_feature    1440..1441
                     /note="zymogen cleavage site"
     misc_feature    1591..1593
                     /note="N-linked glycosylation site"
     misc_feature    1792..1794
                     /note="N-linked glycosylation site"
     misc_feature    2264..2269
                     /note="polyadenylation signal"
     polyA_site      2282
                     /note="polyadenylation site"
BASE COUNT      587 a    661 c    648 g    490 t
ORIGIN

Initial Score    =    160  Optimized Score  =    398  Significance =  9.73
Residue Identity =    51%  Matches          =    487  Mismatches   =   341
Gaps             =    123  Conservative Substitutions             =     0

            X         10        20              30              40        50
            GTCGACT--GGAACGAGACGACC-TGCTCTG----TCA--CAGGCATG-AGGGGTCCCCGGC
            | | |||   |  |  | || ||  ||| ||    |||   || || | |    | || |||
   CTACGTTGGAGGCTACTTCTGTTCCTGCCGTCCAGGCTATGAGCTTCAGGAAGACAGGCATTCCTGCCAGGC
       560 X     570       580       590       600       610       620

        60            70        80        90        100       110
   AGAGATGA-CA--GTGCTG-GCGC-CAGCCTGGAGCCCAAAT-TCCTCCCTGTTGCTGCTGTTGCTGCTGCT
```



10. US-08-162-407-1 (1-879)
HUMC1R Human complement C1r mRNA, complete cds.

LOCUS HUMC1R 2493 bp ss-mRNA PRI 31-OCT-1994
DEFINITION Human complement C1r mRNA, complete cds.
ACCESSION M14058
KEYWORDS .
SOURCE Human, cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE 1 (bases 1 to 2493)
AUTHORS Leytus,S.P., Kurachi,K., Sakariassen,K.S. and Davie,E.W.
TITLE Nucleotide sequence of the cDNA coding for human complement C1r
JOURNAL Biochemistry 25, 4855-4863 (1986)
STANDARD full automatic
COMMENT NCBI gi: 179643
FEATURES Location/Qualifiers
source 1..2493
/organism="Homo sapiens"
mat_peptide 64..1452
/note="C1r A chain"
/codon_start=1
CDS 64..2181
/gene="C1R"
/map="12p13"
/note="human complement C1r; NCBI gi: 179644"
/codon_start=1
/gdb_xref="G00-119-729"
/translation="MWLLYLLVPALFCRAGGSIPIPQKLFGEVTSPLFPKPYPNNFET
TTVITVPTGYRVKLVFQQFDLEPSEGCFYDYVKISADKKSLGRFCGQLGSPLGNPPGK
KEFMSQGNKMLLTFHTDFSNEENGTIMFYKGFLAYYQAVDLDECASRSKSGEEDPQPQ
CQHLCHNYVGGYFCSCRPGYELQEDRHSCQAECSSELYTEASGYISSLEYPRSYPPDL
RCNYSIRVERGLTLHLKFLEPFDIDDHQQVHCPYDQLQIYANGKNIGEFCGKQRPPDL
DTSSNAVDLLFFTDESGDSRGWKLRYTTEIIKCPQPKTLDEFTIIQNLQPQYQFRDYF
IATCKQGYQLIEGNQVLHSFTAVCQDDGTWHRAMPRCKIKDCGQPRNLPNGDFRYTTT
MGVNTYKARIQYYCHEPYYKMQTRAGSRESEQGVYTCTAQGIWKNEQKGEKIPRCLPV
CGKPVNPVEQRQRIIGGQKAKMGNFPWQVFTNIHGRGGGALLGDRWILTAAHTLYPKE
HEAQSNASLDVFLGHTNVEELMKLGNHPIRRVSVHPDYRQDESYNFEGDIALLELENS
VTLGPNLLPICLPDNDTFYDLGLMGYVSGFGVMEEKIAHDLRFVRLPVANPQACENWL
RGKNRMDVFSQNMFCAGHPSLKQDACQGDSGGVFAVRDPNTDRWVATGIVSWGIGCSR
GYGFYTKVLNYVDWIKKEMEEED"
mat_peptide 1453..2178

```
                     /note="C1r B chain serine protease"
                     /codon_start=1
 BASE COUNT      619 a    680 c    662 g    532 t
 ORIGIN      Chromosome 12p13.

Initial Score    =    160  Optimized Score  =    397  Significance =  9.73
Residue Identity =    51%  Matches          =    486  Mismatches   =   342
Gaps             =    123  Conservative Substitutions             =     0

            X         10        20              30        40        50
            GTCGACT--GGAACGAGACGACC-TGCTCTG----TCA--CAGGCATG-AGGGGTCCCCGGC
            | | |||   |  |  | || ||  ||| ||    |||   || || | |    | || |||
    CTACGTTGGAGGCTACTTCTGTTCCTGCCGTCCAGGCTATGAGCTTCAGGAAGACAGGCATTCCTGCCAGGC
        570   X   580       590       600       610       620       630

          60          70        80        90       100       110
    AGAGATGA-CA--GTGCTG-GCGC-CAGCCTGGAGCCCAAAT-TCCTCCCTGTTGCTGCTGTTGCTGCTGCT
     |||   | ||  | ||||  | |  || |   || | | || |||  ||||  | | |  | | | || |
    TGAGTGCAGCAGCGAGCTGTACACGGAGGCATCAGGCTACATCTCCAGCCTGGAG-TACCCTCGGTCCTAC-
      640       650       660       670       680       690       700

   120       130           140       150       160       170       180
    GAGTCCTTGCCTGCGGGG----GACACCTGACTGTT--ACTTCAGCCACAGTCCCATCTCCTCCAACTTCAA
        |||  ||||||  |      ||| |   | | |  |      ||| ||  |||  | ||| ||| |||
    -CCCCCTGACCTGCGCTGCAACTACAGCATCCGGGTGGAGCGGGGCCTCA--CCCTGCACCT-CAAGTTC--
       710       720       730       740       750       760       770

      190       200       210       220       230       240       250
    AGTGAAG--TTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTC----AATCT
      || ||  ||| || | |||| |||||||| ||   |||  ||| | | | || || ||  |     ||||
    -CTGGAGCCTTTTGATA-TTGA-TGACCACCAGC---AAG--TACACTGCCCCTATGACCAGCTACAGATCT
           780       790       800           810       820       830

          260       270       280       290       300       310
    -TCAGGACGAGAAG--CACTG-CAAGGCCT-TGTGGAGCCTCTTCCTAGCCCAGCGCTGGATA-GAGCAACT
     |    ||| ||||  || || | ||  || || | |||     ||   ||| |  || || |  |||| |
    ATGCCAACGGGAAGAACATTGGCGAGTTCTGTGGGAAGCAAAGGCC---CCCCGACCTCGACACCAGCAGC-
      840       850       860       870       880       890       900

     320       330       340       350       360       370       380
    GAAGACTGTGG-CAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCT
     ||  ||||||    | | |       || | |  || || ||||  |      |  ||  |       ||||
    -AATGCTGTGGATCTGCTGTTCTTCACAGATGAGTC-GGGGGACAGCCGGGGCTGGAAGCTGCGCTACACCA
         910       920       930       940       950       960       970

    390       400       410       420       430       440       450
    CATGTACCTTCCA--GCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGA
    |    | | || |  ||||| | |  || | |   | | | ||  || ||||  |||   |||| ||
    CCGAGATCATCAAGTGCCCCCAGCCCAA-GAC--C--CTAGACGAGTTCACCATCATCCAGAACCTGCAG--
        980       990      1000          1010      1020      1030

    460       470       480           490       500       510       520
    CACCTGCACACAGCTGCTTG-CT-CTGAA--GC--CCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTG
    |  | | || ||| | | || || ||   |  ||  |||| |   | |||| || ||||    |     ||
    CCTCAGTAC-CAGTTCCGTGACTACTTCATTGCTACCTGCA--AGCAAGG-CTACCAGCTCATAGAGGGGAA
  1040      1050      1060      1070        1080      1090      1100

       530       540       550       560       570       580       590
    CCTGGAGGTGCAGTGC--CAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGA--GTCCCATAGCCCTAGAAG
    || || | |||| | |  || | |   ||  |||   || || | |   | |  || |||| |||| ||| |
```



```
    CCAGGTGCTGCATTCCTTCA-CAGCTGTCTGCCAGGATGATGGCACGTGGCATCGTGCCAT-GCCC-AGATG
    1110      1120      1130      1140      1150      1160      1170

          600       610       620       630       640       650
    CCACGGAGCTCCCAGAGCCT---CGGCCCAGGCAGCTGTTGCTCCTGCTG-CTGCTGCTGC-CTCTCACACT
    | |  ||  ||   ||   |   | |||| |  | |||   ||  || || || | | | | |   |||| |
    CAA--GA--TCAAGGACTGTGGGCAGCCCCGAAACCTG--CCTAATGGTGACTTCCGTTACACCACCACAAT
       1180      1190      1200      1210      1220      1230      1240

      660       670       680       690       700       710
    G---GTG--CTGCTGGCAG--CCGCCTGGGG-CCTTC-GC--TG-GC-AAAGGGCAAGAAGGAGGGGGGAGC
    |   |||  |  ||   || |||  |   |  || | ||  || || | |   ||||| | ||    ||||
    GGGAGTGAACACCTACAAGGCCCGTATCCAGTACTACTGCCATGAGCCATATTACAAGATGCAGACCAGAGC
         1250      1260      1270      1280      1290      1300      1310

     720       730       740       750       760       770       780
    T-CCACCCTGGGG--TGCCCCTCCCCT---CCCATCCCTAGGATTCGAGCCTTGTGCATCGTTGACTCAGCC
    |  || |  || |  ||  |       |   | | | |  || |   | || || || |    |||  |||
    TGGCAGCAGGGAGTCTGAGCAAGGGGTGTACACCTGCACAGCA-CAGGGCATT-TGGAAGAATGA-ACAG-A
       1320      1330      1340      1350      1360      1370      1380

        790       800       810       820       830       840       850
    AGGGTCTTATCTCGGTTACACCTGTAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGA
    ||||    |    | || | |       |  |||    |||||  || | | |      |  |||  |   | |
    AGGGAGAGA---AGATTCCTCGGTGCTTGCCAGTGTGTGGGAAGCCCGTG-AACCCCGTGGAACAG-AGGCA
         1390      1400      1410      1420      1430      1440

      860       870       X
    GCAGGTCGTC-TCGTTCCAGTCGAC
    ||   |  ||   |  | |   | |
    GCGCATAATCGGAGGGCAAAAAGCCAAGATGGGCA
  1450      1460      1470  X   1480
```

11. US-08-162-407-1 (1-879)
    HUMCEL Human carboxyl ester lipase (CEL) gene, complete c

```
LOCUS       HUMCEL      11502 bp ds-DNA             PRI       01-NOV-1994
DEFINITION  Human carboxyl ester lipase (CEL) gene, complete cds.
ACCESSION   M94579
KEYWORDS    carboxyl ester lipase.
SOURCE      Homo sapiens (individual_isolate 1) (tissue library: lambda-DASH)
            male DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 11502)
  AUTHORS   Lidberg,U., Nilsson,J., Stromberg,K., Stenman,G., Sahlin,P.,
            Enerback,S. and Bjursell,G.
  TITLE     Genomic organization, sequence analysis, and chromosomal
            localization of the human carboxyl ester lipase (CEL) gene and a
            CEL-like (CELL) gene
  JOURNAL   Genomics 13, 630-640 (1992)
  STANDARD  full automatic
COMMENT     NCBI gi: 180243
FEATURES             Location/Qualifiers
     source          1..11502
                     /organism="Homo sapiens"
                     /isolate="1"
                     /cell_type="lymphocyte"
```

```
                /sequenced_mol="DNA"
                /sex="male"
                /tissue_lib="lambda-DASH"
exon            1630..1717
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=1
mRNA            join(1630..1717,4043..4193,4279..4401,4679..4876,
                6160..6290,6468..6575,6718..6835,8301..8487,8687..8890,
                10096..10293,10620..11460)
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
CDS             join(1643..1717,4043..4193,4279..4401,4679..4876,
                6160..6290,6468..6575,6718..6835,8301..8487,8687..8890,
                10096..10293,10620..11364)
                /gene="CEL"
                /map="9q34.3"
                /EC_number="3.1.1.3"
                /note="NCBI gi: 180244"
                /codon_start=1
                /gdb_xref="G00-127-527"
                /product="carboxyl ester lipase"
                /translation="MLTMGRLQLVVLGLTCCWAVASAAKLGAVYTEGGFVEGVNKKLG
                LLGDSVDIFKGIPFAAPTKALENPQPHPGWQGTLKAKNFKKRCLQATITQDSTYGDED
                CLYLNIWVPQGRKQVSRDLPVMIWIYGGAFLMGSGHGANFLNNYLYDGEEIATRGNVI
                VVTFNYRVGPLGFLSTGDANLPGNYGLRDQHMAIAWVKRNIAAFGGDPNNITLFGESA
                GGASVSLQTLSPYNKGLIRRAISQSGVALSPWVIQKNPLFWAKKVAEKVGCPVGDAAR
                MAQCLKVTDPRALTLAYKVPLAGLEYPMLHYVGFVPVIDGDFIPADPINLYANAADID
                YIAGTNNMDGHIFASIDMPAINKGNKKVTEEDFYKLVSEFTITKGLRGAKTTFDVYTE
                SWAQDPSQENKKKTVVDFETDVLFLVPTEIALAQHRANAKSAKTYAYLFSHPSRMPVY
                PKWVGADHADDIQYVFGKPFATPTGYRPQDRTVSKAMIAYWTNFAKTGDPNMGDSAVP
                THWEPYTTENSGYLEITKKMGSSSMKRSLRTNFLRYWTLTYLALPTVTDQEATPVPPT
                GDSEATPVPPTGDSETAPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDS
                GAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAP
                PVPPTGDSGAPPVTPTGDSETAPVPPTGDSGAPPVPPTGDSEAAPVPPTDDSKEAQMP
                AVIRF"
intron          1718..4042
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=1
exon            4043..4193
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=2
intron          4194..4278
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=2
exon            4279..4401
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=3
intron          4402..4678
                /gene="CEL"
```



```
                /map="9q34.3"
                /note="G00-127-527"
                /number=3
exon            4679..4876
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=4
intron          4877..6159
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=4
exon            6160..6290
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=5
intron          6291..6467
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=5
exon            6468..6575
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=6
intron          6576..6717
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=6
exon            6718..6835
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=7
intron          6836..8300
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=7
exon            8301..8487
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=8
intron          8488..8686
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=8
exon            8687..8890
                /gene="CEL"
                /map="9q34.3"
                /note="G00-127-527"
                /number=9
intron          8891..10095
                /gene="CEL"
```

```
                     /map="9q34.3"
                     /note="G00-127-527"
                     /number=9
     exon            10096..10293
                     /gene="CEL"
                     /map="9q34.3"
                     /note="G00-127-527"
                     /number=10
     intron          10294..10619
                     /gene="CEL"
                     /map="9q34.3"
                     /note="G00-127-527"
                     /number=10
     exon            10620..11460
                     /gene="CEL"
                     /map="9q34.3"
                     /note="G00-127-527"
                     /number=11
BASE COUNT     2461 a   3514 c   3264 g   2263 t
ORIGIN

Initial Score    =    156  Optimized Score  =    395  Significance =  9.43
Residue Identity =    51%  Matches          =    492  Mismatches   =   319
Gaps             =    151  Conservative Substitutions               =     0

            X       10        20        30        40        50
            GTCGACTGGAACGAGAC-GACCTG--CTCTGTCACAGGCATGAGGGGTCCCCGGCAGAGATG
            |||||| | |  | ||| ||| ||  |||||  ||||       |||||| ||||
   ACCGGCTGTGGTCGACAGAAGAGGGACAGACGTGGCCTCTG---CAGGTCCACTCGGTCCCTGGCA------
      550    X  560       570       580       590       600

   60        70        80        90       100       110       120
    ACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGTCC---TTGC
          | |||  || |  || | |  ||  |     ||| | || ||| |  |  ||  || ||    ||
    -----CCCGGCCGCA-CGGGGTGGCAGAA--CGGGAGTGTGGTTGGTGTGG--GAAGCACAGGCCCCAGTGT
         610       620       630       640       650       660

  130       140       150       160       170       180       190
    CTGC-GGGGGACACCTGACTGTTA---CTTCAGCCACAGTC-CCATCTCCTCCA--ACTTCAAAGTGAAGTT
    || | |||||||   ||  ||  |   ||   ||  |  || || | | | |||  ||| || || |  ||
    CTCCTGGGGGACTGTTGGGTGGGAAGGCTCTGGCTGCCCTCACCCTGTTC-CCATCACTGCAGAGGGCTGT-
  670       680       690       700       710       720       730

       200       210       220       230       240       250       260
    TAGAGAGTTGACTGACCACCTGCTTAAAGA-TTACCCAGTCACTGTGGCCGTCAATCT-TCAGGACGAGAAG
      | | || | |||   | ||||  |  || |  | | || |  ||| || |||  ||  | |  |  |||
    --GCG-GTGGCCTG--GAGCTGC-CACTGAGTGTCTCGGTGAGGGTGACC-TCACACTGGCTGAGCTTAAAG
    740       750       760       770       780       790       800

             270       280       290       300       310       320
    ----CA-CTGCAAGGCCTT----GTGGAG-CCT----CTTCCTAG-CCCAGCGCTGGATAGAGCAACTGAAG
        || ||| ||| | ||    |||| |  ||    ||||  ||  ||    |||| |  || |  | ||
    GCCCCATCTG-AAGACTTTGTTCGTGGTGTTCTTTCACTTCTCAGAGCCTTTCCTGGCTCCAGGA--TTAAT
       810       820       830       840       850       860       870

             330       340       350       360       370       380
    A-CTG-TGGCAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCA-ACACCGAGATACATTTTGTCACCTCA
    | ||| |  |||    || ||  |     || ||  |  || ||| ||  |    | | ||   || || |
    ACCTGTTCACAGAAAATACGAGTCGCCTCCTCCTCCACAACCTCACACGACCTTCTCCCTTCCCTC-CCGCT
        880       890       900       910       920       930       940
```



```
  390       400       410       420       430       440       450
    TG-TACCTTCCAGCCCCTAC---CAGAATGTCTGCGATTCGTCCAGACCAACATCT---CCC---ACCTCCT
      |   | ||||  ||||| |   ||   || ||| |  |   ||||  |  |  ||   |||      |||
    GGCCTCTTTCCCTCCCCTTCTGTCACTCTGCCTGGGCATGCCCCAGGGCCTCGGCTGGGCCCTTTGTTTCCA
       950       960       970       980       990      1000      1010

            460       470       480       490       500       510       520
    GAAGGACACCTGCACACAGCTGCTTGCTCT-GAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGT
     | ||| |||| ||       || |||  || || |||     || | |   ||  |||      | | | |
    CAGGGAAACCTACA------TGGTTGGGCTAGATGCC-----TCCGCA--CCCCCCCACCCACACCCCCTGA
    1020      1030              1040              1050      1060      1070

          530       540       550       560       570       580       590
    GCCTGGAGGTGCAGTGCCAGCC-GGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCT-AGAAG-
    ||||  | || |  | ||  || ||||   ||   |  | ||    |   ||| ||| |||| | ||  |
    GCCTCTA-GTCC--TCCCTCCCAGGACAGATCAGGCTGGATGGTGAC----AGTTCCACAGCCTTGAGTGGG
     1080      1090      1100      1110              1120      1130      1140

              600       610       620       630       640       650
    -CCAC--GGAGCT-CCCAGAGCCTCGGCCCAGGCAGCTGTTGCT-CCTGCTGCTGCTGCTGCCTCTCACACT
     |  |   | ||| | | | |  |||   ||  ||||     || || ||| |    ||  |     |  ||
    ACTGCCTTGTGCTGCTCTGGGATTCG---CACCCAGCTTGGACTACCCGCTCCACGGGCCCCAGGAAAAGCT
          1150      1160      1170      1180      1190      1200

   660       670       680       690       700       710
    GGTGCTGCT--GG-CAGCCGCCTG------GGGCCTTC-----GCTGGC---AAAGGGCAAGAAG-GAGGGG
     || | | |  || |||| | ||      |||||| |     |||| |      |  | ||||  | | |
    CGTACAGATAAGGTCAGCCACATGAGTGGAGGGCCTGCAGCATGCTGCCCTTTCTGTCCCAGAAGTCACGTG
1210      1220      1230      1240      1250      1260      1270      1280

          720       730       740       750       760       770       780
    GGAGCTCCACCCTGGGGTGCCCCTCCCCTCCCA--TCCCTAGGATTCGAGCCTTGTG-CATCGTTGACTCAG
       | ||  ||||  |  ||||||     || |     | || |   | || | | | ||| |  |||  ||
    CTCGGTC--CCCTCTGAAGCCCCTTTGGACCTAGGGGACAAGCA---GGGCATGGAGACAT-GGAGACAAAG
          1290      1300      1310      1320      1330      1340

          790       800       810       820       830       840
    CCAGGGTCTTATCTCGGTTACACCTGTAATCTCAGCCCTTGGGAGC---CCAGA--GCAGG--ATTGCTGAA
      | |  ||| ||||  | |||  || |  |  ||||| || || |   |||||  |||||  | || || |
    -TATGCCCTTTTCTC--TGACA-GTG-ACACCAAGCCC-TGTGAACAAACCAGAAGGCAGGGCACTG-TGCA
    1350      1360              1370      1380      1390      1400      1410

       850       860       870       X
    ---TGGTCTGGAGCAGGTCGTCTCGTTCCAGTCGAC
       ||  | |   ||        | | | |||  || |
    CCCTGCCCGGCCCCACCATCCCCCTTACCACCCGCCACCTTGCCAC
       1420      1430      1440      1450


12. US-08-162-407-1 (1-879)
    HSCEL        Human carboxyl ester lipase (CEL) gene, complete c

 ID   HSCEL      standard; DNA; PRI; 11502 BP.
 XX
 AC   M94579;
 XX
 DT   15-JUL-1992 (Rel. 32, Created)
 DT   16-DEC-1994 (Rel. 42, Last updated, Version 6)
 XX
```

```
DE   Human carboxyl ester lipase (CEL) gene, complete cds.
XX
KW   carboxyl ester lipase.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
XX
RN   [1]
RP   1-11502
RA   Lidberg U., Nilsson J., Stromberg K., Stenman G., Sahlin P.,
RA   Enerback S., Bjursell G.;
RT   "Genomic organization, sequence analysis, and chromosomal
RT   localization of the human carboxyl ester lipase (CEL) gene and a
RT   CEL-like (CELL) gene";
RL   Genomics 13:630-640(1992).
XX
DR   CPGISLE; HSCEL; Release 2.0.
DR   SWISS-PROT; P19835; BAL_HUMAN.
XX
CC   NCBI gi: 180243
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..11502
FT                   /organism="Homo sapiens"
FT                   /isolate="1"
FT                   /cell_type="lymphocyte"
FT                   /sequenced_mol="DNA"
FT                   /sex="male"
FT                   /tissue_lib="lambda-DASH"
FT   exon            1630..1717
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=1
FT   mRNA            join(1630..1717,4043..4193,4279..4401,4679..4876,6160..629
0
FT                   ,6468..6575,6718..6835,8301..8487,8687..8890,10096..10293,
FT                   10620..11460)
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT   CDS             join(1643..1717,4043..4193,4279..4401,4679..4876,6160..629
0
FT                   ,6468..6575,6718..6835,8301..8487,8687..8890,10096..10293,
FT                   10620..11364)
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /EC_number="3.1.1.3"
FT                   /note="NCBI gi: 180244"
FT                   /codon_start=1
FT                   /gdb_xref="G00-127-527"
FT                   /product="carboxyl ester lipase"
FT   intron          1718..4042
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=1
FT   exon            4043..4193
```



```
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=2
FT   intron          4194..4278
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=2
FT   exon            4279..4401
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=3
FT   intron          4402..4678
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=3
FT   exon            4679..4876
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=4
FT   intron          4877..6159
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=4
FT   exon            6160..6290
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=5
FT   intron          6291..6467
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=5
FT   exon            6468..6575
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=6
FT   intron          6576..6717
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=6
FT   exon            6718..6835
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=7
FT   intron          6836..8300
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=7
FT   exon            8301..8487
```

```
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=8
FT   intron          8488..8686
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=8
FT   exon            8687..8890
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=9
FT   intron          8891..10095
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=9
FT   exon            10096..10293
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=10
FT   intron          10294..10619
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=10
FT   exon            10620..11460
FT                   /gene="CEL"
FT                   /map="9q34.3"
FT                   /note="G00-127-527"
FT                   /number=11
XX
SQ   Sequence 11502 BP; 2461 A; 3514 C; 3264 G; 2263 T; 0 other;
```

```
Initial Score    =    156  Optimized Score  =    395  Significance =  9.43
Residue Identity =    51%  Matches          =    492  Mismatches   =   319
Gaps             =    151  Conservative Substitutions               =     0
```

```
             X       10         20         30         40         50
             GTCGACTGGAACGAGAC-GACCTG--CTCTGTCACAGGCATGAGGGGTCCCCGGCAGAGATG
             |||||| | |  | ||| ||| ||  |||||  ||||       |||||| ||||
   ACCGGCTGTGGTCGACAGAAGAGGGACAGACGTGGCCTCTG---CAGGTCCACTCGGTCCCTGGCA------
      550    X  560        570        580        590        600

  60         70         80         90        100        110        120
   ACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGTCC---TTGC
         | |||  || |  || | |  ||  |     ||| | || ||| |  |  ||  || ||    ||
   -----CCCGGCCGCA-CGGGGTGGCAGAA--CGGGAGTGTGGTTGGTGTGG--GAAGCACAGGCCCCAGTGT
        610        620        630        640        650        660

  130        140        150        160        170        180        190
   CTGC-GGGGGACACCTGACTGTTA---CTTCAGCCACAGTC-CCATCTCCTCCA--ACTTCAAAGTGAAGTT
   || | |||||||   ||  ||  |   ||   ||  |  || || | | | |||  ||| || || |  ||
   CTCCTGGGGGACTGTTGGGTGGGAAGGCTCTGGCTGCCCTCACCCTGTTC-CCATCACTGCAGAGGGCTGT-
 670        680        690        700        710        720        730

      200        210        220        230        240        250        260
```



```
   TAGAGAGTTGACTGACCACCTGCTTAAAGA-TTACCCAGTCACTGTGGCCGTCAATCT-TCAGGACGAGAAG
     | | || | |||   | ||||  |  || |  | | || |  ||| || |||  ||  | |  |  |||
   --GCG-GTGGCCTG--GAGCTGC-CACTGAGTGTCTCGGTGAGGGTGACC-TCACACTGGCTGAGCTTAAAG
   740        750        760        770        780        790        800

            270        280        290        300        310        320
   ----CA-CTGCAAGGCCTT----GTGGAG-CCT----CTTCCTAG-CCCAGCGCTGGATAGAGCAACTGAAG
       || ||| ||| | ||    |||| |  ||    ||||  ||  ||    |||| |  || |  | ||
   GCCCCATCTG-AAGACTTTGTTCGTGGTGTTCTTTCACTTCTCAGAGCCTTTCCTGGCTCCAGGA--TTAAT
      810        820        830        840        850        860        870

            330        340        350        360        370        380
   A-CTG-TGGCAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCA-ACACCGAGATACATTTTGTCACCTCA
   | ||| |  |||    || ||  |    || ||  |  || ||| ||  |    | | ||   || || |
   ACCTGTTCACAGAAAATACGAGTCGCCTCCTCCTCCACAACCTCACACGACCTTCTCCCTTCCCTC-CCGCT
       880        890        900        910        920        930        940

 390        400        410        420        430        440        450
   TG-TACCTTCCAGCCCCTAC---CAGAATGTCTGCGATTCGTCCAGACCAACATCT---CCC---ACCTCCT
     |   | ||||  ||||| |   ||   || ||| |  |   ||||  |  |  ||   |||       |||
   GGCCTCTTTCCCTCCCCTTCTGTCACTCTGCCTGGGCATGCCCCAGGGCCTCGGCTGGGCCCTTTGTTTCCA
      950        960        970        980        990       1000       1010

        460        470        480        490        500        510        520
   GAAGGACACCTGCACACAGCTGCTTGCTCT-GAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGT
    | ||| |||| ||       || |||  || || |||     || | |   ||  |||       | | | |
   CAGGGAAACCTACA------TGGTTGGGCTAGATGCC-----TCCGCA--CCCCCCCACCCACACCCCCTGA
   1020       1030           1040           1050       1060       1070

        530        540        550        560        570        580        590
   GCCTGGAGGTGCAGTGCCAGCC-GGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCT-AGAAG-
   ||||  | || |  | ||  || ||||   ||   |  | ||    |   ||| ||| |||| | ||  |
   GCCTCTA-GTCC--TCCCTCCCAGGACAGATCAGGCTGGATGGTGAC-----AGTTCCACAGCCTTGAGTGGG
    1080       1090       1100       1110           1120       1130       1140

            600        610        620        630        640        650
   -CCAC--GGAGCT-CCCAGAGCCTCGGCCCAGGCAGCTGTTGCT-CCTGCTGCTGCTGCTGCCTCTCACACT
    |  |   | ||| | | | |  |||   ||  |||||    || || ||| |    ||  |     | ||
   ACTGCCTTGTGCTGCTCTGGGATTCG---CACCCAGCTTGGACTACCCGCTCCACGGGCCCCAGGAAAAGCT
         1150       1160       1170       1180       1190       1200

 660        670        680        690        700        710
   GGTGCTGCT--GG-CAGCCGCCTG------GGGCCTTC-----GCTGGC---AAAGGGCAAGAAG-GAGGGG
    || | | |  || ||||| | ||      ||||||  |    |||| |       |  | |||||  | | |
   CGTACAGATAAGGTCAGCCACATGAGTGGAGGGCCTGCAGCATGCTGCCCTTTCTGTCCCAGAAGTCACGTG
1210       1220       1230       1240       1250       1260       1270       1280

        720        730        740        750        760        770        780
   GGAGCTCCACCCTGGGGTGCCCCTCCCCTCCCA--TCCCTAGGATTCGAGCCTTGTG-CATCGTTGACTCAG
     | ||  ||||  |  ||||||     || |     | || |   | || | | | ||| |  ||| ||
   CTCGGTC--CCCTCTGAAGCCCCTTTGGACCTAGGGGACAAGCA---GGGCATGGAGACAT-GGAGACAAAG
         1290       1300       1310       1320           1330       1340

        790        800        810        820        830        840
   CCAGGGTCTTATCTCGGTTACACCTGTAATCTCAGCCCTTGGGAGC---CCAGA--GCAGG--ATTGCTGAA
     |  ||| |||| | |||  || |  |  ||||| || || |   |||||  |||||  | || || |
   -TATGCCCTTTTCTC--TGACA-GTG-ACACCAAGCCC-TGTGAACAAACCAGAAGGCAGGGCACTG-TGCA
   1350       1360           1370       1380       1390       1400       1410

      850        860        870        X
```

```
    ---TGGTCTGGAGCAGGTCGTCTCGTTCCAGTCGAC
       ||  | |   ||      | | | |||  || |
    CCCTGCCCGGCCCCACCATCCCCCTTACCACCCGCCACCTTGCCAC
        1420      1430      1440      1450
```

13. US-08-162-407-1 (1-879)
    DDU03413 Dictyostelium discoideum AX2 calcium binding prote

```
LOCUS       DDU03413     1679 bp    mRNA            INV       21-NOV-1993
DEFINITION  Dictyostelium discoideum AX2 calcium binding protein mRNA, complete
            cds.
ACCESSION   U03413
KEYWORDS    .
SOURCE      slime mold.
  ORGANISM  Dictyostelium discoideum
            Eucaryotae; Protozoa; Sarcomastigophora; Sarcodina; Rhizopoda;
            Eumycetozoa; Dictyostelia; Dictyosteliida; Dictyostelium.
REFERENCE   1  (sites)
  AUTHORS   Wennington,R., Greenwood,M. and Tsang,A.
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1679)
  AUTHORS   Wennington,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1993) Wennington R., Concordia University, 1455
            De Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8, Canada
  STANDARD  full automatic
COMMENT     NCBI gi: 416312
FEATURES             Location/Qualifiers
     source          1..1679
                     /organism="Dictyostelium discoideum"
                     /strain="AX2"
                     /dev_stage="vegetative"
     CDS             124..1527
                     /note="NCBI gi: 416313"
                     /codon_start=1
                     /product="calcium binding protein"
                     /translation="MQNPQNPPPAGSAADFYSQMPVKVMGTPGAPGSQSTPGAPGAPG
                     QYPPQQPGAPGSNLPPYPGTQQPGAPGAPGQYPPQQPGQYPPQQPGAPGQYPPQQPGQ
                     PGYPPQQPGQSGQYPPQQPGQPGYPPQQPGAPGQYPPQQGQPGQYPPQQPGQPGQYPP
                     QQQGQYPPQQPGQPGAYPPQQPGQPGAYPPQQGVQNTLAKTGAPGQPGVPPPQGAYPG
                     QPGVPPQQGAYPGQQPPMGAYPPQGQPGAYPPQQQQVAYPGQQPPMGAY
                     PPQQGAYPGQQGAYPGQQGAYPGQQGAYPGQPPMGAYPGQQQYGVSAYSQTTAAYQTG
                     AYPGQTPMGVYPGQTTAYAAYQTTAIGYGATSTYSYTRPTYSYATPYSRTAAFVVPVG
                     LPPHVVQKMMAASAAFRIHDSNCSGTLSKKEFKKLIKHLGYYFSKGQTKMLFHSIDRD
                     YSGSLSEREFVDWWSMQ"
     misc_feature    1357..1395
                     /note="calcium binding site"
BASE COUNT      567 a    429 c    251 g    432 t
ORIGIN
```

```
Initial Score    =    155  Optimized Score  =    369  Significance =  9.36
Residue Identity =    48%  Matches          =    463  Mismatches   =   349
Gaps             =    140  Conservative Substitutions               =     0

            X       10        20        30            40            50
            GTCGACTGGAACGAGACGACCTGCTCTGTCA--CAGG-CA-TGAGGGG----TCCCCGGCA-
            ||| ||    | | | | |||  |     ||  |||| ||   ||| |    ||| |  ||
   CAACAACCAGGTCAAC---CAGGTGCCTACCCACCACAACAACCAGGTCAACCAGGTGCCTATCCACCACAA
```



```
630       640           650       660       670       680       690

       60          70          80          90          100         110         120
-GAGATG-AC--AGTGCTGGCGCC-AGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGA
  || || ||  | | |    ||| || | || || | |  ||  |       | | |||| | |  | |
CAAGGTGTACAAAATACACTTGCCAAGACAGGTGCTCCAGGTCAAC-------CAGGTGTTCCACCACC--A
 700       710       720       730       740             750

       130       140       150       160         170         180
GTCCTTGCCTGCGGGGGACACCTGACTGTTACTTCAGCCACA--GTCCCATCTCC---TC--CAACTTCAAA
     |||||     || || |    |||| |  || | |||  || || | |||   ||  ||||  | ||
 CAAGGTGCCTATCCAGGTCAACCAGGTGTTCCACCA-CAACAAGGTGCC-TATCCAGGTCAACAACCACCAA
760       770       780       790       800       810       820

   190       200       210       220       230       240         250
GTGAAGTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCA--ATCTTCAGG
  |  | |||   |   |  | | ||| |    || ||| |||  |||   | |   ||  | |||
 TGGGTGCTTATCCACCACAAGGTCAACCAG----GAGCTTATCCA-CCACAAGGTCAACCAGGAGCTTATCC
830       840       850           860       870       880       890

   260       270         280       290       300       310       320
ACGAGAAGCACTGCAAG--GCCT-TGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTG
|| | ||  ||  |||| || | |   | | |  ||   || |  | ||  ||   || |  || |  |
ACCACAACAACAACAAGTAGCATATCCAGGTCAACAACC--ACCAATGGGTGCTTA-TCCACCACAACAAGG
 900       910       920       930           940       950       960

   330       340       350       360       370         380       390
TGGCAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTT---TGTCACCTCATGTA
| |||   ||  || |  | | | | ||     | |||||||   || | ||| |    ||||  |  | ||
T-GCA-TATC-CAGGTCAACAAGGTGCTTATCCAGGTCAACA---AGGTGCATATCCAGGTCAAC-AAGGTG
    970       980       990      1000        1010      1020      1030

     400       410       420       430       440       450       460
CCT-TCCAGCCCCTACCA-GAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGA-CACCT
| | |||||   | ||||  ||||  ||| ||  |||||  |||||    | |  |  || | |  || |
CATATCCAG-GTCAACCACCAATGGGTGC-AT--ATCCAGGTCAACA---ACAATATGGTGTATCAGCATAT
       1040      1050      1060      1070        1080      1090

         470       480       490             500       510       520
GCAC-----ACAGCTGCTTGCTCTGAAGCCCTG--TATC--GG--GAAGGCCTGCCAGAATTTCTCTCGGT-
  |||     || |||||||  ||  ||   ||||  ||  ||  ||     |  ||| ||  |||
TCACAAACTACCGCTGCTT-ATC-AAACTGGTGCTTATCCAGGTCAAACACCAATGGGTGTTTATCCAGGTC
 1100      1110        1120      1130      1140      1150      1160

         530       540       550             560       570       580
----GCCTGGAGGTGCAGTGCCAGCCGGACTCCT-CCACC----CTGCTGC--CCCCAAGGAGTCCCATAGC
      ||| |  |||  ||||   |  ||  || | | |      || |||  |  |||     |  | || |
AAACTACTGCATATGC--TGCCTATCAAACAACTGCTATCGGTTATGGTGCAACATCAACTTAT-TCTTA-C
 1170      1180        1190      1200      1210      1220      1230

    590       600       610       620           630       640       650
CCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGC-AG--CTGTTGCTCCTGCTGCTGCTGCTGCCTCT
 |||| | | ||  | |||  |  ||  |  |  |  | ||  | | |||| ||  | | | | ||  |
ACTAG-ACCAACTTA-CTCTTA-TGCAACTCCATATTCTAGAACCGCTGCTTTTGTAGTTCCAGTTGGTTTA
    1240      1250      1260      1270      1280      1290      1300

     660       670       680       690           700       710       720
CACACTGGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTGGCAAAGG--GCAAGAAGGAGGGGGGAGCTCCAC
|   |   || ||   ||          ||||      |||| |  |||   || |     ||  ||||
CCACCACATGTTG--TACAAAAAATGATGGCC-------GCAAGTGCTGCATTCAGAATTCATGA-TTCCA-
```

```
      1310        1320       1330              1340       1350       1360

       730        740        750        760        770        780        790
CCTGGGGTGCCCCTCCCCTCCCATCCCTAGGATTCGAGCCTTGTGCATCGTTGACTCAGCCAGGGT-CT-TA
  ||  |||   ||    ||  |    || |||  |           || || |  ||   ||| | || |
ATTGTAGTGGTACT-TTATCTAA--AAAAGAATTTAA-----AAAACTCATT-AAACATTTAGGTTACTATT
   1370       1380       1390           1400       1410       1420

          800        810        820        830        840        850
TCTC----GGTTACACCTGTAATCTCAGCCCTTGGGA--GCCC-AGA--GCAG-GATTGCTGAA-TGGTCTG
||||    ||| |  | ||  ||| | |  || |   |  |  |  ||   ||| | ||  | || ||  |
TCTCTAAAGGTCA-AACTAAAATGTTATTCCATTCAATTGATCGTGATTACAGTGGTTCATTAAGTGAACGT
1430       1440       1450       1460       1470       1480       1490

    860        870        X
GAGCAGGTCGTCTCGTTCCAGTCGAC
||    ||  |  ||  |  |   ||   |
GAATTTGTTGACT-GGTGGAGTATGCAATAAATTAA
1500       1510       1520 X     1530
```

14. US-08-162-407-1 (1-879)
    HSNUMAU4G H.sapiens NuMA gene (Clone U4).

```
LOCUS       HSNUMAU4G    2832 bp    RNA             PRI       15-APR-1993
DEFINITION  H.sapiens NuMA gene (Clone U4).
ACCESSION   Z14228
KEYWORDS    alternative splicing; NuMA gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
            Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 2832)
  AUTHORS   Tang,T.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-AUG-1992) to the EMBL/GenBank/DDBJ databases. Tang T.
            K., Institute of Biomedical Sciences, Laboratory of Cell Biology,
            Section 2, Academy Road, Taipei, Taiwan, Republic of China, 115
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 2832)
  AUTHORS   Tang,T.K., Tang,C.J., Chen,Y.L. and Wu,C.W.
  TITLE     Nuclear proteins of the bovine esophageal epithelium. II. The NuMA
            gene gives rise to multiple mRNAs and gene products reactive with
            monoclonal antibody W1
  JOURNAL   J. Cell. Sci 104, 249-260 (1993)
  STANDARD  full automatic
COMMENT     NCBI gi: 296119
FEATURES             Location/Qualifiers
     source          1..2832
                     /organism="Homo sapiens"
                     /cell_line="U251"
                     /clone_lib="U251 cDNA library"
                     /clone="Clone U4"
     mRNA            <1..2409
                     /gene="NuMA"
                     /note="alternative splice product"
                     /evidence=experimental
BASE COUNT      690 a    724 c    987 g    431 t
ORIGIN
```



```
Initial Score    =    153  Optimized Score  =    403  Significance =  9.20
Residue Identity =    51%  Matches          =    491  Mismatches   =   334
Gaps             =    128  Conservative Substitutions               =     0

            X         10         20         30         40         50
            GTC-GACTGGAACG-AG--ACGA-CCTGCTCTG---TCACAGGCATGAGGGGTCCCCGGCAG
            | | | |||||||| ||   ||| ||| | |   || | ||| |||| ||  |   ||
    GGAGCAGGCTGACAGCCTGGAACGCAGCCTCGAGGCTGAGCGGGCCTCCCGGGC-TGAGCGGGACAGTGCTC
    500        510        520        530        540        550        560

        60         70         80         90        100        110        120
    AGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGT-CCT
      | |||  ||| || |||||  | |||   ||       ||| |  |||   |   | |  | |||| |||
    TGGAGACTCTGCAGG-GCCAG-TTAGAGGAGAA----GGCCCAGGAGCT-AGGGCACAG--TCAGAGTGCCT
  570        580        590        600        610        620        630

       130        140        150        160        170        180        190
    T-GCCTGCGGGGGACACCTGACT--GTT-ACTTCAGC-CACAGTCCCATCTCCTCCAACTTCAAAG-TGAAG
    | |||| |||   | ||  || |  | |  |||| || |  ||   ||   ||    || | ||||| |||||
    TAGCCT-CGGCCCA-ACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAAGACC----ACAGCAAGGCTGAAG
           640        650        660        670        680        690

           200        210        220        230        240        250
    TTTAGAGAGTTGAC-----TGACC-----ACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCA--
      | || |    | |    || ||      || ||   | | | |   ||  |   | ||| || ||| ||
    ATGAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCTGA--GAGGAAAAATAGCC-TC-ATCAGCAGC
    700        710        720        730        740        750        760

           260        270        280        290        300        310
    --GGACGAGAAGCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAA----CTGAA
      ||| ||| ||   | ||   || ||     ||| |||| |   || |  ||  |||||||      ||||
    TTGGAGGAGGAGGTGTCCA--TCC-TGAATCGCCAGGTCCTGGAGAAG-GAGGGGGAGAGCAAGGAGTTGAA
      770        780        790        800        810        820        830

    320          330        340        350        360        370        380
    --GACTG----TGGCAGGGTCTAAGATGCAAACGCTTCTGGAGGA-CGTCAACACCGAGATACATTTTGTCA
      | |||     |||| | |||  ||| |      |  ||||||||  | |  | ||  | | ||     |  |
    GCGGCTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTGGAGGAGAGGCTGCGCC-TGCTGCAGGCAGAGA
       840        850        860        870        880        890        900

       390        400        410          420        430        440        450
    CCTCATGTACC--TTCCAGCCCCTACCAGAATG----TCTGCGATTCGTCCAGACCAACATCTCCCACCTCC
    |  |  | | |  | ||||   || | |||| |    |||||  | ||    ||     ||        |||||
    CAGCCAGCAACAGTGCCAG-AGCTGC-AGAACGCAGCTCTGCTCTGCGGGAGGAGGTGCA----GAGCCTCC
       910        920        930        940        950        960          970

           460        470        480        490        500        510        520
    TGAAGGACACCTGCACACAGCTGCTTGC-TCTGAAGCCCTGTATCGGGAAGGCCTG-CCAGAATTTCTCTCG
      | ||||  |    | ||||| | | || || ||   ||||   ||| | |  ||| ||  |     ||
    GGGAGGAGGCTGAGAAACAGCGGGTGGCTTCAGAGAACCTG---CGGCAGGAGCTGACCTCACAGGCTGAGC
           980        990       1000       1010         1020       1030

           530        540          550            560        570        580
    GTGCCTGGAGGTGCAGTGCC---AGCCGGACTCCT-CCA----CCCTGCTGCCCCCAAGGAGTCCCATAGCC
    ||||  ||||| || | |||    |  | |  | |  ||        | ||  ||  || |||    ||  |||
    GTGC--GGAGGAGCTGGGCCAAGAATTGAAGGCGTGGCAGGAGAAGTTCT-TCCAGAAAGAG---CA-GGCC
  1040       1050       1060       1070       1080       1090         1100

       590        600          610        620        630        640          650
    CTAGAAGCCACGGAGCTC--CCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTG-CTGCT--GCCTC
```

```
CTCTCCACCCTGCAGCTCGAGCACA-CCAGCACACAGGC-CCTGGTGAGTGAGCTGCTGCCAGCTAAGCACC
  1110      1120       1130      1140       1150      1160      1170

      660       670         680       690       700           710
TCACACTGGTGCTGCTGGC--AGCCGCCTGGGGCCTTCGCTGGCAAAGGGCA--AGAAGGAG--GGGGGAGC
TCTGCCAGCAGCTGCAGGCCGAGCAGGCCGCTGCCGAGAAACGCCACCGTGAGGAGCTGGAGCAGAGCAAGC
  1180      1190      1200      1210      1220      1230      1240

720          730         740         750       760       770       780
TCCACCCTGGGG---TGC-CCCTCCCCTCC--CATCCCTAGGATTCGAGCCTTGTGCATC-GTTGACTCAGC
AGGCCGCTGGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAG--CGGGAG-CTTGGGGAGCTGATTCCTCTGC
1250      1260      1270      1280      1290      1300      1310

       790       800       810       820       830       840       850
--CAGGGTCTTATCTCGGTTACACCTGTAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCT
GGCA-GAAGGTGGCAGAGCAGGAGCGAACAGCTCAGCAGCTGCGGG--CAGAGAAGGCCAGCT--ATG--CA
  1320      1330      1340      1350      1360      1370      1380

     860       870       X
GGAGCAGGTCGTCTCGTTCCAGTCGAC
G-AGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTG
       1390      1400      1410
```

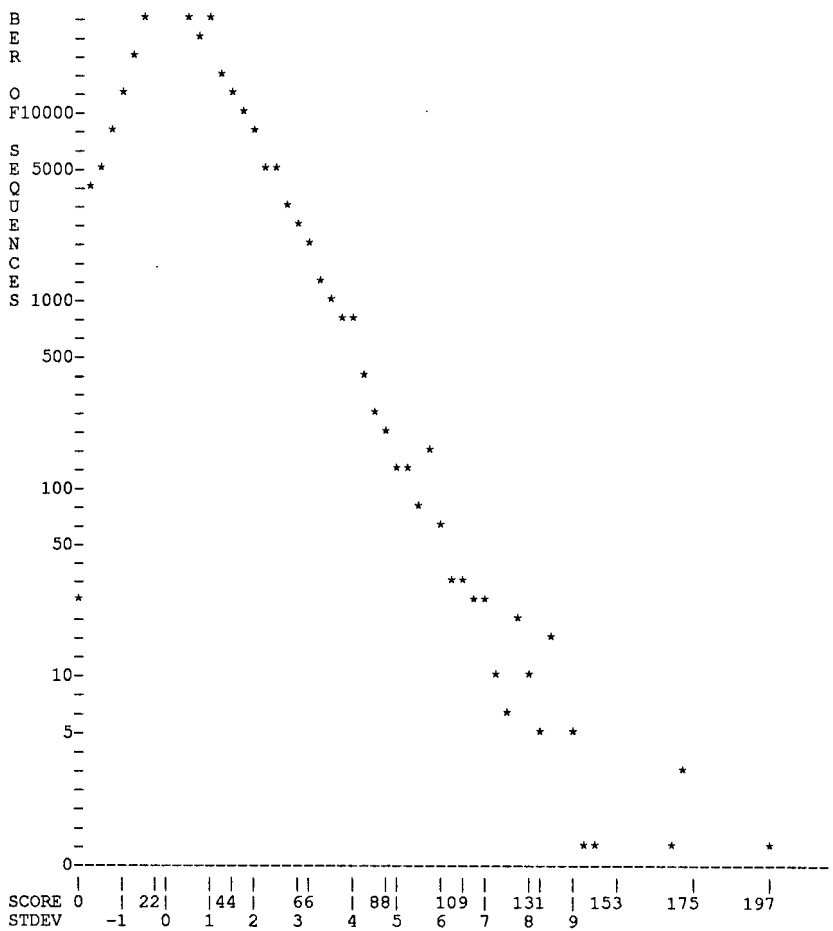
15. US-08-162-407-1 (1-879)
    HSNUMAU6G H.sapiens NuMA gene (Clone U6).

```
LOCUS       HSNUMAU6G    4762 bp    RNA             PRI       15-APR-1993
DEFINITION  H.sapiens NuMA gene (Clone U6).
ACCESSION   Z14227
KEYWORDS    alternative splicing; NuMA gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
            Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 4762)
  AUTHORS   Tang,T.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-AUG-1992) to the EMBL/GenBank/DDBJ databases. Tang T.
            K., Institute of Biomedical Sciences, Laboratory of Cell Biology,
            Section 2, Academy Road, Taipei, Taiwan, Republic of China, 115
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 4762)
  AUTHORS   Tang,T.K., Tang,C.J., Chen,Y.L. and Wu,C.W.
  TITLE     Nuclear proteins of the bovine esophageal epithelium. II. The NuMA
            gene gives rise to multiple mRNAs and gene products reactive with
            monoclonal antibody W1
  JOURNAL   J. Cell. Sci 104, 249-260 (1993)
  STANDARD  full automatic
COMMENT     NCBI gi: 296120
FEATURES             Location/Qualifiers
     source          1..4762
                     /organism="Homo sapiens"
                     /cell_line="U251"
                     /clone_lib="U251 cDNA library"
```



```
                     /clone="Clone U6"
     mRNA            <1..3465
                     /gene="NuMA"
                     /note="alternative splice product"
                     /evidence=experimental
BASE COUNT     1231 a    1180 c    1579 g     772 t
ORIGIN

Initial Score    =    153  Optimized Score  =    403  Significance =  9.20
Residue Identity =    51%  Matches          =    491  Mismatches   =   334
Gaps             =    128  Conservative Substitutions              =     0

            X        10          20          30        40        50
            GTC-GACTGGAACG-AG--ACGA-CCTGCTCTG---TCACAGGCATGAGGGGTCCCCGGCAG
GGAGCAGGCTGACAGCCTGGAACGCAGCCTCGAGGCTGAGCGGGCCTCCCGGGC-TGAGCGGGACAGTGCTC
    1600  X   1610      1620      1630      1640      1650      1660

     60        70        80        90        100       110       120
AGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGT-CCT
TGGAGACTCTGCAGG-GCCAG-TTAGAGGAGAA----GGCCCAGGAGCT-AGGGCACAG--TCAGAGTGCCT
   1670      1680      1690      1700      1710      1720

   130      140         150         160       170       180       190
T-GCCTGCGGGGGACACCTGACT--GTT-ACTTCAGC-CACAGTCCCATCTCCTCCAACTTCAAAG-TGAAG
TAGCCT-CGGCCCA-ACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAAGACC-----ACAGCAAGGCTGAAG
1730        1740      1750      1760      1770      1780      1790

      200         210         220       230       240       250
TTTAGAGAGTTGAC-----TGACC----ACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCA--
ATGAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCTGA--GAGGAAAAATAGCC-TC-ATCAGCAGC
    1800      1810      1820      1830      1840      1850      1860

      260       270       280       290       300       310
--GGACGAGAAGCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAA----CTGAA
TTGGAGGAGGAGGTGTCCA--TCC-TGAATCGCCAGGTCCTGGAGAAG-GAGGGGGAGAGCAAGGAGTTGAA
      1870      1880      1890      1900      1910      1920

320         330       340       350       360       370       380
--GACTG----TGGCAGGGTCTAAGATGCAAACGCTTCTGGAGGA-CGTCAACACCGAGATACATTTTGTCA
GCGGCTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTGGAGGAGAGGCTGCGCC-TGCTGCAGGCAGAGA
1930      1940      1950      1960      1970      1980      1990

     390       400       410       420       430       440       450
CCTCATGTACC--TTCCAGCCCCTACCAGAATG----TCTGCGATTCGTCCAGACCAACATCTCCCACCTCC
CAGCCAGCAACAGTGCCAG-AGCTGC-AGAACGCAGCTCTGCTCTGCGGGAGGAGGTGCA----GAGCCTCC
2000      2010      2020      2030      2040      2050      2060

       460       470       480       490       500       510       520
TGAAGGACACCTGCACACAGCTGCTTGC-TCTGAAGCCCTGTATCGGGAAGGCCTG-CCAGAATTTCTCTCG
GGGAGGAGGCTGAGAAACAGCGGGTGGCTTCAGAGAACCTG---CGGCAGGAGCTGACCTCACAGGCTGAGC
 2070      2080      2090      2100      2110      2120      2130
```

```
          530       540         550         560       570       580
    GTGCCTGGAGGTGCAGTGCC---AGCCGGACTCCT-CCA----CCCTGCTGCCCCCAAGGAGTCCCATAGCC
    |||| ||||| || | |||   |   | |  | |  ||       | ||  ||  || |||   ||  |||
    GTGC--GGAGGAGCTGGGCCAAGAATTGAAGGCGTGGCAGGAGAAGTTCT-TCCAGAAAGAG---CA-GGCC
        2140      2150      2160      2170      2180      2190

       590       600         610       620       630       640         650
    CTAGAAGCCACGGAGCTC--CCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTG-CTGCT--GCCTC
    ||     ||  | |||||   || | ||    | |||||  ||| ||     |||||||| | |||  ||  |
    CTCTCCACCCTGCAGCTCGAGCACA-CCAGCACACAGGC-CCTGGTGAGTGAGCTGCTGCCAGCTAAGCACC
 2200      2210      2220      2230      2240      2250      2260

         660       670         680       690       700         710
    TCACACTGGTGCTGCTGGC--AGCCGCCTGGGGCCTTCGCTGGCAAAGGGCA--AGAAGGAG--GGGGGAGC
    ||   | |  ||||| |||  ||| | | |  |||       || |  |  |  ||  ||||  | |  |||
    TCTGCCAGCAGCTGCAGGCCGAGCAGGCCGCTGCCGAGAAACGCCACCGTGAGGAGCTGGAGCAGAGCAAGC
2270      2280      2290      2300      2310      2320      2330      2340

    720         730       740         750       760       770       780
    TCCACCCTGGGG---TGC-CCCTCCCCTCC--CATCCCTAGGATTCGAGCCTTGTGCATC-GTTGACTCAGC
       | |||||||   |||   |    || |  |   || ||     ||| |||| | | | | |  ||| ||
    AGGCCGCTGGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAG--CGGGAG-CTTGGGGAGCTGATTCCTCTGC
       2350      2360      2370      2380      2390      2400      2410

         790       800       810       820       830       840       850
    --CAGGGTCTTATCTCGGTTACACCTGTAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCT
      || |    |  |   |    | |    | ||||||   || | |  ||||| |||   |||  |||  |
    GGCA-GAAGGTGGCAGAGCAGGAGCGAACAGCTCAGCAGCTGCGGG--CAGAGAAGGCCAGCT--ATG--CA
         2420      2430      2440      2450      2460      2470

        860       870       X
    GGAGCAGGTCGTCTCGTTCCAGTCGAC
    | |||||  |   |  | |  ||  | |
    G-AGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTG
      2480      2490      2500      2510
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq1inv.res made by  on Wed 5 Apr 95 21:32:49-PDT.


Query sequence being compared:US-08-162-407-1' (1-879)
Number of sequences searched:               313646
Number of scores above cutoff:                4947

      Results of the initial comparison of US-08-162-407-1' (1-879) with:
   Data bank : EMBL-NEW 1, all entries
   Data bank : GenBank 86, all entries
   Data bank : GenBank-NEW 1, all entries
   Data bank : UEMBL 41_86, all entries

100000-
      -
N     -
U50000-
M     -      ***
```

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		

Initial scores to save 45 Alignments to save 15
Optimized scores to save 0 Display context 10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	30	30	12.72

Times:	CPU	Total Elapsed
	01:48:40.01	01:48:46.00

Number of residues: 288682119
Number of sequences searched: 313646
Number of scores above cutoff: 4947

Cut-off raised to 23.
Cut-off raised to 27.
Cut-off raised to 30.
Cut-off raised to 33.
Cut-off raised to 36.
Cut-off raised to 39.
Cut-off raised to 41.
Cut-off raised to 44.
Cut-off raised to 47.
Cut-off raised to 50.
Cut-off raised to 53.
Cut-off raised to 56.
Cut-off raised to 59.
Cut-off raised to 62.
Cut-off raised to 65.
Cut-off raised to 67.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 13 standard deviations above mean ****					
1. GCRINVA	G.crassicaudatus involucrin g	1155	197	406	13.13	0
	**** 11 standard deviations above mean ****					
2. ATRINVSA	Owl monkey involucrin (small	894	172	347	11.16	0
3. HUMINV2	Human involucrin gene, exon 2	2108	172	414	11.16	0
4. HUMINV2	Human involucrin gene, exon 2	2108	172	414	11.16	0
	**** 10 standard deviations above mean ****					
5. CEZK638	Caenorhabditis elegans cosmid	1762	167	403	10.77	0
	**** 9 standard deviations above mean ****					
6. DMU13397	Drosophila melanogaster Orego	2091	145	406	9.04	0
	**** 8 standard deviations above mean ****					
7. CRUPA1V	Chinese hamster pro-alpha-1 (	6114	144	408	8.96	0
8. GORINVOLUB	Gorilla gorilla involucrin ge	1818	141	407	8.73	0
9. HUMCA1V	Human mRNA for collagen alpha	5676	139	397	8.57	0
10. HUMPA1V	Human pro-alpha-1 (V) collage	7138	139	399	8.57	0
11. HUMPA1V	Human pro-alpha-1 (V) collage	7138	139	399	8.57	0
12. HS4ULIR3	epstein-barr virus simple rep	1150	134	370	8.18	0

13. CHPINVOL	Chimpanzee (P.paniscus) invol	1683	134	407	8.18	0
14. U02454	Cloning vector pCMVEBNA, comp	5452	134	385	8.18	0
15. U02428	Cloning vector pDR2, complete	10737	134	385	8.18	0
16. U02455	Cloning vector rpDR2, complet	10850	134	385	8.18	0
17. EBV	Epstein-Barr virus (EBV) geno	172281	134	385	8.18	0
18. HS4B958RAJ	Epstein-Barr virus, artifactu	184113	134	385	8.18	0
19. HS4B958RAJ	Epstein-Barr virus, artifactu	184113	134	385	8.18	0
20. EBV	Epstein-Barr virus (EBV) geno	172281	134	385	8.18	0
21. EBV	Epstein-Barr virus (EBV) geno	172281	134	385	8.18	0
22. HUMARC1	Human androgen receptor gene,	1753	133	399	8.10	0
23. SRCEL1G	S.reticuli cell gene for myce	2760	133	391	8.10	0
24. HUMANDREC	Human androgen receptor (AR)	3569	133	399	8.10	0
25. I09510	Sequence 1 from patent WO 890	3569	133	399	8.10	0
26. HSANDREC	Human androgen receptor (AR)	3569	133	399	8.10	0
27. I08490	Sequence 3 from patent WO 870	2000	132	390	8.02	0
28. MMSCSLN8B	M.musculus rearranged T-cell	4560	132	377	8.02	0
	**** 7 standard deviations above mean ****					
29. HSKER65A	Human DNA for 65 kD keratin t	1192	131	295	7.94	0
30. DRO3AKLP	Drosophila melanogaster (chro	4394	130	394	7.86	0
31. USTNFR20S3	Mus musculus tumor necrosis f	1481	129	200	7.78	0
32. GIBINVOL	Hylobates lar involucrin gene	1569	129	405	7.78	0
33. MUSAGRIN3	Mus musculus agrin gene, exon	1952	127	387	7.63	0
34. ORAINVOL	Orang-utan involucrin gene, c	2508	127	408	7.63	0
35. HUMIGCB1	Human Ig germline J-mu-delta	3182	127	376	7.63	0
36. HUMIGCB1	Human Ig germline J-mu-delta	3182	127	376	7.63	0
37. HSIGCB1	Human Ig germline J-mu-delta	3182	127	376	7.63	0
38. CGPD2	Cricetulus griseus nucleotide	13664	127	386	7.63	0
39. HUMIGLZC	Human immunoglobulin light ch	482	126	216	7.55	0
40. HUMIGLZB	Human immunoglobulin light ch	482	126	217	7.55	0
41. XAADHLB	Xanthobacter autotrophicus ha	1015	126	373	7.55	0
42. PFACSNURI	Plasmodium knowlesi circumspo	1485	126	371	7.55	0
43. HSCOL4A2	Human mRNA for alpha-2 chain	3416	126	385	7.55	0
44. M78027	EST01614 Homo sapiens cDNA cl	416	125	187	7.47	0
45. HUMIGLZD	Human immunoglobulin light ch	482	125	216	7.47	0

1. US-08-162-407-1' (1-879)
GCRINVA G.crassicaudatus involucrin gene, complete cds.

LOCUS GCRINVA 1155 bp ds-DNA PRI 15-SEP-1990
DEFINITION G.crassicaudatus involucrin gene, complete cds.
ACCESSION J05437
KEYWORDS involucrin.
SOURCE G.crassicaudatus fibroblast DNA.
ORGANISM Galago crassicaudatus
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Primates; Strepsirhini; Galagidae.
REFERENCE 1 (bases 1 to 1155)
AUTHORS Phillips,M., Djian,P. and Green,H.
TITLE The involucrin gene of the Galago: Existence of a correction process acting on its segment of repeats
JOURNAL J. Biol. Chem. 265, 7804-7807 (1990)
STANDARD full automatic
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted by P.Dijian, 26-MAR-1990.

NCBI gi: 176996
FEATURES Location/Qualifiers
source 1..1155
/organism="Galago crassicaudatus"

```
     CDS             1..1155
                     /note="involucrin;  NCBI gi: 176997"
                     /codon_start=1
                     /translation="MSQQHTLPVTLPPALRQELLGTLSPPAAAQQEQRKQPTPLPTAC
                     QKVGSELPGEVPSKHEEKGTDPVKGVLEQECGQQEPELHLGKQQDVHLMKRQDPQEPE
                     LHLGKQPEPEGPEPHLGKEQQHQESQDPELHLGKQQQQESQEQELYPGKQQEPQDPEL
                     HLGKQQQQESQEQGLCLIKQREPQESQEQRLHLGKEQESQEQRLHLGEEQASQEQRLH
                     LGEEQASQEQRLHLGEEQASPEQRLQLLPQGPQEQELHLGKQQQQQESQQHQEQHEEH
                     QKAEDLGQQHRQEKAQREQQLEEQLDEGKKLLDQQLDQEAVKRHEQLQRDEQFGMKKE
                     QLLEPPGQQKGQLEKPVFVPVPGQVQDIQPAQTAKGEALLLPEQPQEPEV"
BASE COUNT      339 a    299 c    379 g    138 t
ORIGIN

Initial Score    =    197  Optimized Score  =    406  Significance = 13.13
Residue Identity =    51%  Matches          =    492  Mismatches   =   334
Gaps             =    121  Conservative Substitutions               =     0

            X       10        20        30        40        50
            GTCGACTGGAACG-AGACGACCTGCTCCAGACCAT----TCAGCAAT-CCTGCTCTGGGCTC
            | | |||  |  |   ||  | ||  |||||   |    || |   | ||||    || | |
   AAAGGAAGCAGCCAACTCCACTGCCCACTGCGTG--CCAGAAGGTGGGCTCCGAGCTCCCTGGAGAGGTCCC
   100       110       120       130       140       150       160

      60        70        80        90       100       110       120
   C-CAAGGGCTGAGATTACAGGTGTA-ACCGAGATAAGACCCTGGCTGAGTCAACGAT----GCACAAG--GC
   | ||| |  ||||   | |||   | | |  |    ||  |||| ||| ||   |||  ||  ||
   CTCAAAGCATGAGGAGAAAGGCACAGATCCTGTAAAAGGGGTGCTTGAG-CAAGAATGTGGGCAGCAGGAGC
   170       180       190       200       210       220       230

             130       140       150       160       170       180
   TCGA----ATCCTAGGGATGGGAG-GGGAGGGGCACCCCAGG--GTGGAGCTCCCCCCTCCTTCTTGCCCTT
    |||      ||| ||||  | ||  ||| | ||| |  | |  | ||     ||| |     |  |  ||
   CCGAGCTGCACCT-GGGAAAGCAGCAGGATGTGCATCTGATGAAGCGGCAGGACCCACAAGAGCCCGAGCTA
  240       250       260       270       280       290       300

     190       200           210       220       230       240       250
   TGCCAGCGAAGGC-CCCAG-GC---GGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCA
     || | |||   |  |||| ||     | ||| | | |||   || |  ||| ||  |||||||||| ||||||
   CACCTGGGAAAACAACCAGAGCCAGAGGGGCCGGAACCACATCTGGGCAAGG-AGCAGCAGCATCAGGAGTC
 310       320       330       340       350       360       370       380

        260       270       280       290       300       310       320
   ACAGCTGCCTGGGCCGAGGCTCTGGGAGCTCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAGGG
   |||    || | ||   |  |||||||   | |  ||    || |   ||| ||  ||   |||| ||
   ACAAGACCCGGAGC--TGCATCTGGGAAAGCAGCAGC----AG--C---AGGAGTCACAGG----AGCA-GG
          390       400       410                 420       430

      330       340       350       360       370       380       390
   TGGAGGAGTCCGGCTGGCA-CTGCACCTCCAGGCACCGAGAG-AAATTCTG-GCAGGCCTTCCCGATACAGG
       | |  | ||    || | | | | |||||  ||| ||||      ||| ||| ||    | |   ||||
   AACTGTATCCTGGAAAACAGCAGGAGCCCCAGG-ACCCAGAGCTGCACCTGGGCAAGC--AGCAGCAGCAGG
    440       450       460       470       480       490       500

         400       410       420       430       440       450       460
   -GCTTCA-GAGCAAGCAGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGGTCTGGACGAATCGCAG
    |  ||| ||||| | | |||||| | |    |    |||  || |    | |   ||||  | ||   ||||
   AGTCTCAGGAGCAGGGA-CTGTGT-CTG--ATAAAACAGCGGGAGCCTCAGG--AGTCT-CAGGA---GCAG
     510       520       530       540       550       560

         470       480       490       500       510       520
```



```
   ACA-TTCTGGTAGGGGCTG-GAAGGTACATGAGGTG-ACAAAATGTATCTCGGTGTTGA--CGTCCTC-CAG
   | | | |   | |||    | | |||     | ||| | | | || | || || |  ||    | | || |||
   AGACTGCACCTGGGGAAAGAGCAGGAGTCTCAGGAGCAGAGACTGCACCTGGGGGAAGAGCAGGCGTCTCAG
   570       580       590       600       610       620       630

 530       540       550       560       570       580       590
   AAGC---GTTTGCATCT-TAGACCCTGCCACAGTCTTCAGTTGCTCTATCCAGC-GCTGGGCTAGGAAGAGG
     |||   |  |||| ||   |     ||   ||| |||| ||   |  | ||  |||||  | ||  |||
   GAGCAGAGACTGCACCTGGGGGAAGAGCAGGCGTC-TCAGGAGC-AGAGACTGCACCTGGGGGAAGAGCAGG
 640       650       660       670       680       690       700

       600       610       620       630       640       650
   ---CTCC--ACAAGGCCTTGCAG-TGCTTC----TCGTCCTGAAGATTGACGGCCACAGTGACTGGGTAATC
      ||||  |  ||    |||||  |||| |     | ||  | ||  |  ||  || |   |||| || |
   CGTCTCCGGAGCAGAGACTGCAGCTGCTGCCCACAGGGACCACAGGA--GCAGG-AACTGCACCTGGGAAAAC
 710       720       730       740       750       760       770

 660       670       680       690       700       710       720
   TTTA--AGCAGGTGGTCAGTCAAC-TCTCTA--AACTTCACTTTGAA-GTTGGA--GGAGATGGGACTGTGG
      |  |||||  ||    |||||  | | |  |||  ||    |||  |  ||  | |||  |  ||| ||
   AGCAGCAGCAGCAGGAGTCTCAACAGCACCAGGAACAGCATGAGGAACATCAGAAAGCAGA-AGACCTGGGG
 780       790       800       810       820       830       840

       730       740       750       760       770       780       790
   CTGAAGTAACAGTCAGGTGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGGAATTT
   | | ||  | || |||| |      |||   ||||  |||||||  | |  ||||||     ||| |||
   CAGCAG-CATAGGCAGGAG---AAAGCA-CAAAGGGAACAGCAGCTGGAGGAGCAGC----TGGATGAA---
 850       860       870       880       890       900

     800       810       820       830       840       850       860
   GGGCTCCAG-GCTGGCGCCAGC-ACTGTCATC-TCTGCCGGGGACCCCTCATG--CCTGTG-ACAGAGCAGG
   |||    ||  || |  ||||| |||| |||    | | | ||     |||| |   || | ||||  |
   GGGAAGAAGCTCTTGGACCAGCAACTG-GATCAAGAGGCAGTGAAGAGACATGAACAACTGCAGAGAGATGA
 910       920       930       940       950       960       970       980

           870       X
   TCGTCTCG--TTCCAGTCGAC
    |   |||   |  ||  |
   ACAATTCGGAATGAAGAAGGAGCAGTTGTTG
         990      1000      1010
```

2. US-08-162-407-1' (1-879)
   ATRINVSA Owl monkey involucrin (small allele) gene, partial

```
LOCUS       ATRINVSA      894 bp ds-DNA             PRI       15-SEP-1990
DEFINITION  Owl monkey involucrin (small allele) gene, partial cds.
ACCESSION   M25314
KEYWORDS    involucrin.
SOURCE      Owl monkey keratinocyte DNA.
  ORGANISM  Aotus trivirgatus
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Platyrrhini; Cebidae; Aotinae.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   Tseng,H. and Green,H.
  TITLE     The involucrin gene of the owl monkey: Origin of the early region
  JOURNAL   Mol. Biol. Evol. 6, 460-468 (1989)
  STANDARD  full automatic
COMMENT     Draft entry and computer-readable sequence for [Mol. Biol. Evol.
            (1989) In press] kindly provided
```

```
            by H.Tseng, 03-JUN-1989.
            The involucrin gene is made up of about 30 base pair type A or B
            repeats of about 30 base pairs.

            NCBI gi: 176559
 FEATURES             Location/Qualifiers
      source          1..894
                      /organism="Aotus trivirgatus"
      CDS             <1..>894
                      /note="involucrin (small allele);  NCBI gi: 176560"
                      /codon_start=1
                      /translation="EFPEQQEGQLKRLEQQEGHLELPEQQEGQLKHLEQQEGQLKHLE
                      QEEKQSELPEQQRGQLKYLEQQEGQLKHLEQQEKQSELPEQQRGQLKYLEQQEGQLKH
                      LEEQKGQLKHLEHQEGQLELPEQVGQPKHLEQLEKQLEHPEQQEGQLKQLEEQEGQVK
                      HLEQQEEQLKHLEQQEGQPKHLEQLEKQLEHLEQQEGQLKHLEQREEQLELPEQQVGQ
                      SKHLEQEEKQLEHPEQQEGQLKHLGKQEAQLELPEQVGQPKHLEQQEKQLEHPEQQEG
                      QLKPQEQQEGQLKGLEQQERQL"
 BASE COUNT       266 a    209 c    350 g     69 t
 ORIGIN

Initial Score    =    172  Optimized Score  =    347  Significance = 11.16
Residue Identity =    52%  Matches          =    419  Mismatches   =   274
Gaps             =    103  Conservative Substitutions               =     0

          130       140       150       160       170       180       190
   TCGAATCCTAGGGATGGGAGGGGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCCCTTTGCCAGC
              | | |   |   |||  |||    |||| | ||||     | |||    ||      || ||
              GAGTTCCCA---GAGCAGCA--GGAGGG-GCAGCT-GAAACGCCT-GGAGC----AGCAAGA
              X          10        20        30        40                 50

        200       210       220       230       240       250       260
   GAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCAACAGCTGCCTGG
   |  ||| || ||  ||| |||| |||| |||   ||| ||||||  | |||| | ||||||  |||  |
   G-GGGCACCTGG-AGCTCCCAG-AGCAGCAG---GAGGGGCAGCTGAAGCACCTGGAGCAGCAGGAGGGGCA
             60        70        80        90       100       110

      270       280       290       300       310       320
   GCCGAGGCTCTGGGA----GCTCCGTGGC-TTCT-AGGGCTATGGGACTCCTTGGGGGCAGCAG-GGT----
   || || || |  |||    |    |  ||  ||| ||  |   | |   |   |||||||||| |  ||
   GCTGAAGCACCTGGAACAGGAGGAGAAGCAGTCTGAGCTCCCAGAGCAGCAGAGGGGGCAGCTGAAGTACCT
    120       130       140       150       160       170       180

      330       340       350       360       370       380       390
   GGAGGAGTCCGGCTGGCA-CTGCACCTCCAGGCACCGAGAGAAATTCTGGCAGGCCTTCCCGATACAGGGCT
   |||| ||   |   |||| ||| | | || || || |   || |  | | | |  || || ||  |||
   GGAGCAGCAGGAGGGGCAGCTGAAGCACCTGGAACAGCAGGAGAAGCAGTCTGAGCTCCCAGA-GCAG----
  190       200       210       220       230       240       250

    400       410       420       430       440       450       460
   TCAGAGCAAGCAGCTG-TGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTG-GTCTGGACGAATCGCAGACA
    |||||   |||||||  || |  | |  |  |||||||  || || ||  |   ||||| ||   ||||| |
   -CAGAGGGGGCAGCTGAAGTACCTG-GAGCAGCAGGAGG-GGCAGCTGAAGCACCTGGAAGA---GCAGAAA
      260       270       280       290       300       310       320

    470       480       490       500       510       520       530
   -TTCTG--GTAGGGGCTGGAAGGTACATGAGGTGACAAAATGTATCTCGGTGTTGACGTCCTCCAG---AAG
      | |  | ||   |||| ||   || |||| | ||   || | |||   |   | ||    |||   |||
   GGGCAGCTGAAGCATCTGG-AGCACCAGGAGG-GGCA-GCTGGAGCTCCCAGAGCAGGTGGGGCAGCCAAAG
         330       340       350       360       370       380       390
```



```
          540       550       560       570       580       590
   CGTTTG---CATCT-TAGACCCTGCCACAGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAG--GAAGAGGCT
   |   ||   || ||  |||  |  | | || |  | |  ||   | | || |  |||| ||  |||| |||
   CACCTGGAACAGCTGGAGAAGCAACTAGAG-CACCCGGAGC---AGC-AG-GAGGGGC-AGCTGAAGCAGCT
         400       410       420       430               440       450

     600       610       620       630       640       650       660
   ---CCACAAGGCCTTGCA-GTGCTTCTCGTCCTGAAGATTGACGGCCACAGTGACTGGGTAATCTTTA--AG
         |  |||     ||| |||   | | |   | ||   ||  |   ||||   ||     ||||  |  ||
   GGAGGAGCAGGAGGGGCAGGTGAAGCACCTGGAGCAGCAGGA--GGAGCAGT---TGAAGCATCTGGAGCAG
     460       470       480       490       500           510       520

       670       680       690       700       710       720       730
   CAGG-TGGTCAGTCAA--CTCTCTAAACTTCACTTTGAAG---TTGGAGGAGATGGGACTGTGGCTGAAGTA
   ||||  || ||| |||  | |    |||  |    ||||    | ||| |  |||     |  || | ||
   CAGGAGGGACAGCCAAAGCACCTGGAACAGC-TGGAGAAGCAACTAGAGCACCTGG---AGCAGCAGGAGGG
        530       540       550       560       570           580       590

           740       750       760       770       780       790
   ACAG-TCAGGTGTCCCCCGCAG-GCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGG--AATTTGGGC
    ||| | | |  ||     |||| || ||||  |||| | |||  |  ||||  |||| |||  | |    ||
   GCAGCTAAAGCATCTGGAGCAGCGCGAGGA-ACAGCTGGAGCTCCCAGAGCAGCAGGTAGGGCAGTCAAAGC
          600       610       620       630       640       650       660

  800       810       820       830       840       850       860
   TCCAG--GCTGGCG--CCAGCACTGTCATCTCTGCCGGGGACCCCTCATGCCTGTGACAG-AGCAGGTCGTC
    || |  || || |    |||| |   | | |  || | |   |  || |   | | ||| | ||  | |
   ACCTGGAGCAGGAGGAGAAGCAGTTGGAACAC--CCAGAG---CAGCAAG--AGGGGCAGCTGAAGCACCTG
         670       680       690           700       710       720

    870       X
   TCGTTCCAGTCGAC
     |   |||  | |
   GGGAAGCAGGAGGCACAGCTGGAG
    730       740       750
```

3. US-08-162-407-1' (1-879)
   HUMINV2 Human involucrin gene, exon 2.

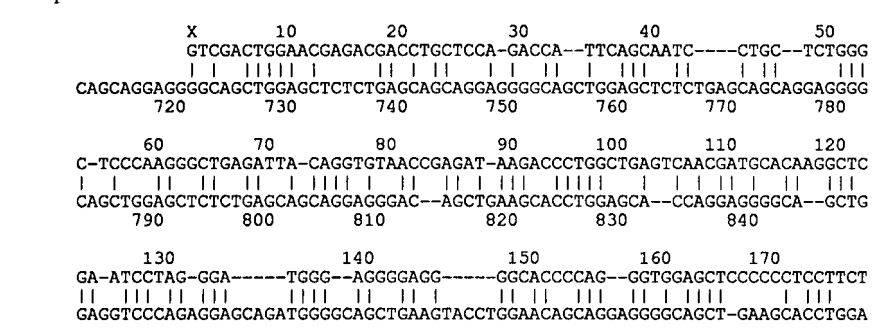
```
 LOCUS       HUMINV2      2108 bp ds-DNA             PRI       06-JAN-1995
 DEFINITION  Human involucrin gene, exon 2.
 ACCESSION   M13903
 KEYWORDS    involucrin; keratinocyte protein.
 SEGMENT     2 of 2
 SOURCE      Human keratinocyte, cDNA to mRNA; and DNA, clone lambda-1-3.
   ORGANISM  Homo sapiens
             Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
             Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
 REFERENCE   1  (bases 1 to 2108)
   AUTHORS   Eckert,R.L. and Green,H.
   TITLE     Structure and evolution of the human involucrin gene
   JOURNAL   Cell 46, 583-589 (1986)
   STANDARD  full automatic
 COMMENT     NCBI gi: 186519
 FEATURES             Location/Qualifiers
      source          1..2108
                      /organism="Homo sapiens"
                      /sequenced_mol="DNA"
      intron          <1..10
```

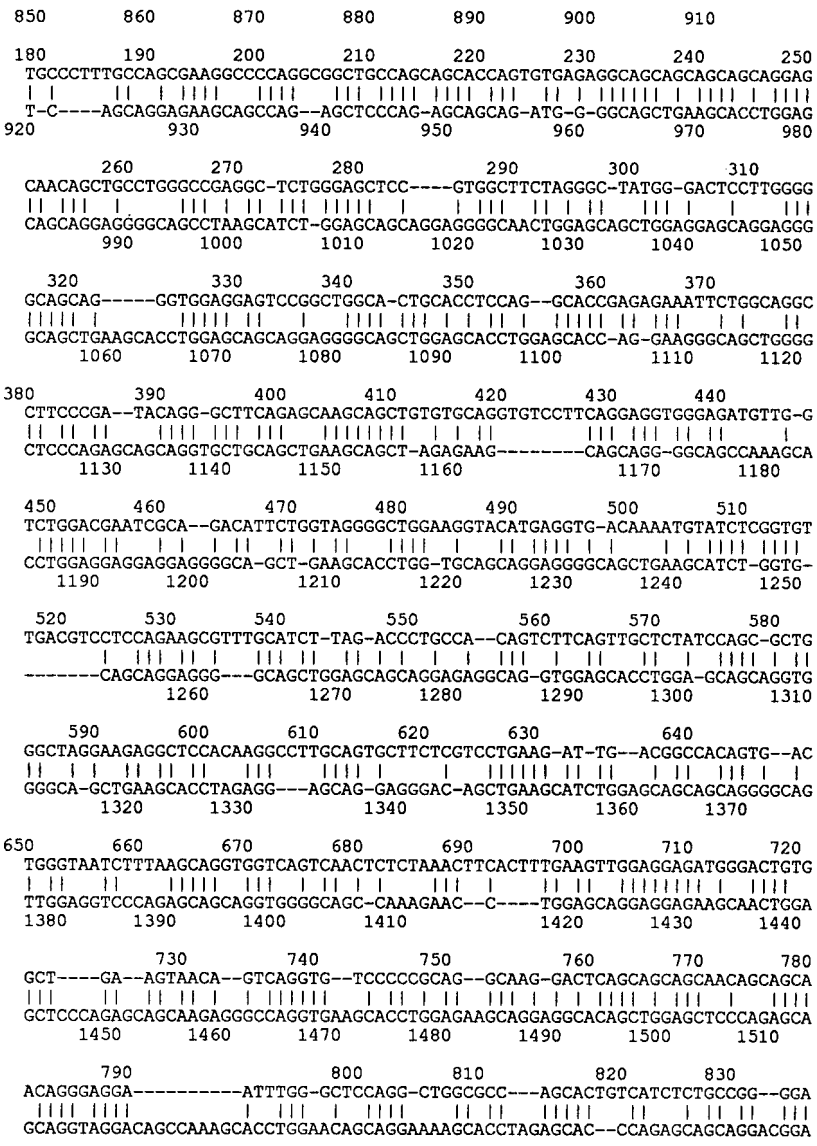
```
                     /gene="IVL"
                     /map="1q21-q22"
                     /note="G00-119-355"
                     /number=1
     CDS             30..1787
                     /gene="IVL"
                     /map="1q21-q22"
                     /note="NCBI gi: 386834"
                     /codon_start=1
                     /gdb_xref="G00-119-355"
                     /product="involucrin"
                     /translation="MSQQHTLPVTLSPALSQELLKTVPPPVNTHQEQMKQPTPLPPPC
                     QKVPVELPVEVPSKQEEKHMTAVKGLPEQECEQQQKEPQEQELQQQHWEQHEEYQKAE
                     NPEQQLKQEKTQRDQQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELPEQQE
                     GHLKHLEQQEGQLKHPEQQEGQLELPEQQEGQLELPEQQEGQLELPEQQEGQLELPEQ
                     QEGQLELPQQQEGQLELSEQQEGQLELSEQQEGQLELSEQQEGQLKHLEHQEGQLEVP
                     EEQMGQLKYLEQQEGQLKHLDQQEKQPELPEQQMGQLKHLEQQEGQPKHLEQQEGQLE
                     QLEEQEGQLKHLEQQEGQLEHLEHQEGQLGLPEQQVLQLKQLEKQQGQPKHLEEEEGQ
                     LKHLVQQEGQLKHLVQQEGQLEQQERQVEHLEQQVGQLKHLEEQEGQLKHLEQQQGQL
                     EVPEQQVGQPKNLEQEEKQLELPEQQEGQVKHLEKQEAQLELPEQQVGQPKHLEQQEK
                     HLEHPEQQDGQLKHLEQQEGQLKDLEQQKGQLEQPVFAPAPGQVQDIQPALPTKGEVL
                     LPVEHQQQKQEVQWPPKHK"
  BASE COUNT      602 a    526 c    711 g    269 t
  ORIGIN      About 1188 bp after segment 1.

Initial Score    =    172  Optimized Score  =    414  Significance = 11.16
Residue Identity =    53%  Matches          =    517  Mismatches   =   297
Gaps             =    156  Conservative Substitutions              =     0

             X       10        20        30        40        50
             GTCGACTGGAACGAGACGACCTGCTCCA-GACCA--TTCAGCAATC-----CTGC--TCTGGG
             | |  ||||| |     || | ||   | |  ||  |  |||  ||     | ||     |||
   CAGCAGGAGGGGCAGCTGGAGCTCTCTGAGCAGCAGGAGGGGCAGCTGGAGCTCTCTGAGCAGCAGGAGGGG
         720       730       740       750       760       770       780

        60        70        80        90       100       110       120
   C-TCCCAAGGGCTGAGATTA-CAGGTGTAACCGAGAT-AAGACCCTGGCTGAGTCAACGATGCACAAGGCTC
   |  |   ||  ||  ||  | |||| |  ||  || | |||  |||||   |  | | || |  ||  |||
   CAGCTGGAGCTCTCTGAGCAGCAGGAGGGAC--AGCTGAAGCACCTGGAGCA--CCAGGAGGGGCA--GCTG
        790       800       810       820       830       840

       130               140               150       160       170
   GA-ATCCTAG-GGA-----TGGG--AGGGGAGG-----GGCACCCCAG--GGTGGAGCTCCCCCCTCCTTCT
   ||  ||| || |||     ||||  ||  || |     || ||  |||  || | ||||     | |||
   GAGGTCCCAGAGGAGCAGATGGGGCAGCTGAAGTACCTGGAACAGCAGGAGGGGCAGCT-GAAGCACCTGGA
  850       860       870       880       890       900       910

  180       190       200       210       220       230       240       250
   TGCCCTTTGCCAGCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAG
   | |     ||  | ||||   ||||   ||| |||| |||| |||  || | ||||||  | |||| | ||||
   T-C----AGCAGGAGAAGCAGCCAG--AGCTCCCAG-AGCAGCAG-ATG-G-GGCAGCTGAAGCACCTGGAG
 920           930       940       950       960       970       980

        260       270       280           290       300       310
   CAACAGCTGCCTGGGCCGAGGC-TCTGGGAGCTCC----GTGGCTTCTAGGGC-TATGG-GACTCCTTGGGG
   || |||  |     ||| | || ||| |||||  |    | |||  || | ||  |||  |    |    |||
   CAGCAGGAGGGGCAGCCTAAGCATCT-GGAGCAGCAGGAGGGGCAACTGGAGCAGCTGGAGGAGCAGGAGGG
         990      1000      1010      1020      1030      1040      1050

     320               330       340       350       360       370
```



```
   GCAGCAG-----GGTGGAGGAGTCCGGCTGGCA-CTGCACCTCCAG--GCACCGAGAGAAATTCTGGCAGGC
   ||||| |       |||||  ||   |  |||| ||| | | || |  |||||  || |||   | |   ||
   GCAGCTGAAGCACCTGGAGCAGCAGGAGGGGCAGCTGGAGCACCTGGAGCACC-AG-GAAGGGCAGCTGGGG
        1060      1070      1080      1090      1100      1110      1120

 380       390       400       410       420       430       440
   CTTCCCGA--TACAGG-GCTTCAGAGCAAGCAGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTG-G
   || || ||    |||| ||| |||   |||||||| | | ||         ||| ||| || ||     |
   CTCCCAGAGCAGCAGGTGCTGCAGCTGAAGCAGCT-AGAGAAG---------CAGCAGG-GGCAGCCAAAGCA
        1130      1140      1150      1160              1170      1180

   450       460       470       480       490       500       510
   TCTGGACGAATCGCA--GACATTCTGGTAGGGGCTGGAAGGTACATGAGGTG-ACAAAATGTATCTCGGTGT
    ||||| ||   | |  | ||  || | ||   |||| |   || |||| | |    | | |||| ||||
   CCTGGAGGAGGAGGAGGGGCA-GCT-GAAGCACCTGG-TGCAGCAGGAGGGGCAGCTGAAGCATCT-GGTG-
      1190      1200      1210      1220      1230      1240      1250

    520       530       540       550       560       570       580
   TGACGTCCTCCAGAAGCGTTTGCATCT-TAG-ACCCTGCCA--CAGTCTTCAGTTGCTCTATCCAGC-GCTG
          | ||| || |   ||| ||  || | |  | |  | ||| |  ||   ||  |  |||| | ||
   -------CAGCAGGAGGG---GCAGCTGGAGCAGCAGGAGAGGCAG-GTGGAGCACCTGGA-GCAGCAGGTG
             1260      1270      1280      1290      1300      1310

       590       600       610       620       630       640
   GGCTAGGAAGAGGCTCCACAAGGCCTTGCAGTGCTTCTCGTCCTGAAG-AT-TG--ACGGCCACAGTG--AC
   ||  | |  || || ||   |||    |||| |      |  |||||| || ||   | ||  ||| |  |
   GGGCA-GCTGAAGCACCTAGAGG---AGCAG-GAGGGAC-AGCTGAAGCATCTGGAGCAGCAGCAGGGGCAG
          1320      1330          1340      1350      1360      1370

 650       660       670       680       690       700       710       720
   TGGGTAATCTTTAAGCAGGTGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGACTGTG
   | ||   ||    |||||  |||  | || | |    |||  |    || ||  |||||||| |  ||||
   TTGGAGGTCCCAGAGCAGCAGGTGGGGCAGC-CAAAGAAC--C----TGGAGCAGGAGGAGAAGCAACTGGA
   1380      1390      1400      1410              1420      1430      1440

             730       740       750       760       770       780
   GCT----GA--AGTAACA--GTCAGGTG--TCCCCCGCAG--GCAAG-GACTCAGCAGCAGCAACAGCAGCA
   |||    ||  || || |  | ||||||   | || | ||  ||| | | | |||| | |||   |  ||||
   GCTCCCAGAGCAGCAAGAGGGCCAGGTGAAGCACCTGGAGAAGCAGGAGGCACAGCTGGAGCTCCCAGAGCA
        1450      1460      1470      1480      1490      1500      1510

        790               800       810       820       830
   ACAGGGAGGA----------ATTTGG-GCTCCAGG-CTGGCGCC---AGCACTGTCATCTCTGCCGG--GGA
    |||| ||||          |  |||  |  ||||    || ||   |||||   ||   | || ||  |||
   GCAGGTAGGACAGCCAAAGCACCTGGAACAGCAGGAAAAGCACCTAGAGCAC--CCAGAGCAGCAGGACGGA
      1520      1530      1540      1550      1560      1570      1580

     840       850       860       870       X
   CCCCTCATGCCTGTGACAGAGCAGGTCGTCTCGTTCCAGTCGAC
   |  || |  | | ||    ||||||  |    | |  ||  |||
   CAACTAAAACATCTGGAGCAGCAGGAGGGGCAGCTGAAG--GACCTGGAGCAGC
      1590      1600      1610      1620      X 1630


4. US-08-162-407-1' (1-879)
   HUMINV2      Human involucrin gene, exon 2.

 LOCUS       HUMINV2      2108 bp ds-DNA             PRI       30-MAR-1994
 DEFINITION  Human involucrin gene, exon 2.
 ACCESSION   M13903
```

```
KEYWORDS    involucrin; keratinocyte protein.
SEGMENT     2 of 2
SOURCE      Human keratinocyte, cDNA to mRNA; and DNA, clone lambda-1-3.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 2108)
  AUTHORS   Eckert,R.L. and Green,H.
  TITLE     Structure and evolution of the human involucrin gene
  JOURNAL   Cell 46, 583-589 (1986)
  STANDARD  full automatic
COMMENT     NCBI gi: 186519
FEATURES             Location/Qualifiers
     source          1..2108
                     /organism="Homo sapiens"
                     /sequenced_mol="DNA"
     intron          <1..10
                     /gene="IVL"
                     /map="1q21-q22"
                     /number=1
     CDS             30..1787
                     /gene="IVL"
                     /map="1q21-q22"
                     /note="NCBI gi: 386834"
                     /codon_start=1
                     /product="involucrin"
                     /translation="MSQQHTLPVTLSPALSQELLKTVPPPVNTHQEQMKQPTPLPPPC
                     QKVPVELPVEVPSKQEEKHMTAVKGLPEQECEQQQKEPQEQELQQQHWEQHEEYQKAE
                     NPEQQLKQEKTQRDQQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELPEQQE
                     GHLKHLEQQEGQLKHPEQQEGQLELPEQQEGQLELPEQQEGQLELPEQQEGQLELPEQ
                     QEGQLELPQQQEGQLELSEQQEGQLELSEQQEGQLELSEQQEGQLKHLEHQEGQLEVP
                     EEQMGQLKYLEQQEGQLKHLDQQEKQPELPEQQMGQLKHLEQQEGQPKHLEQQEGQLE
                     QLEEQEGQLKHLEQQEGQLEHLEHQEGQLGLPEQQVLQLKQLEKQQGQPKHLEEEEGQ
                     LKHLVQQEGQLKHLVQQEGQLEQQERQVEHLEQQVGQLKHLEEQEGQLKHLEQQQGQL
                     EVPEQQVGQPKNLEQEEKQLELPEQQEGQVKHLEKQEAQLELPEQQVGQPKHLEQQEK
                     HLEHPEQQDGQLKHLEQQEGQLKDLEQQKGQLEQPVFAPAPGQVQDIQPALPTKGEVL
                     LPVEHQQQKQEVQWPPKHK"
BASE COUNT      602 a    526 c    711 g    269 t
ORIGIN      About 1188 bp after segment 1.

Initial Score    =    172  Optimized Score  =    414  Significance = 11.16
Residue Identity =    53%  Matches          =    517  Mismatches   =   297
Gaps             =    156  Conservative Substitutions               =     0
```





```
      1520       1530       1540       1550       1560        1570       1580

       840        850        860        870        X
    CCCCTCATGCCTGTGACAGAGCAGGTCGTCTCGTTCCAGTCGAC
    |  || |  | | ||    ||||||  |    | |  ||  |||
    CAACTAAAACATCTGGAGCAGCAGGAGGGGCAGCTGAAG--GACCTGGAGCAGC
       1590       1600       1610       1620       X 1630
```

5. US-08-162-407-1' (1-879)
   CEZK638 Caenorhabditis elegans cosmid ZK638.

```
LOCUS       CEZK638      1762 bp    DNA             INV       14-OCT-1994
DEFINITION  Caenorhabditis elegans cosmid ZK638.
ACCESSION   Z12018
KEYWORDS    .
SOURCE      nematode.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Animalia; Eumetazoa; Nematoda; Secernentea; Rhabditia;
            Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae.
REFERENCE   1  (bases 1 to 1762)
  AUTHORS   Metzstein,M. and Hawkins,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1992) jes@mrc-lmba.cambridge.ac.uk or
            rw@nematode.wustl.edu
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1762)
  AUTHORS   Sulston,J., Du,Z., Thomas,K., Wilson,R., Hillier,L., Staden,R.,
            Halloran,N., Green,P., Thierry-Mieg,J., Qiu,L., Dear,S.,
            Coulson,A., Craxton,M., Durbin,R., Berks,M., Metzstein,M.,
            Hawkins,T., Ainscough,R. and Waterston,R.
  TITLE     The C. elegans genome sequencing project: a beginning [see
            comments]
  JOURNAL   Nature 356, 37-41 (1992)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 1762)
  AUTHORS   Roberts,L.
  TITLE     The Worm Project
  JOURNAL   Science 248, 1310-1313 (1990)
  STANDARD  full automatic
COMMENT     NOTES:
            This is a cosmid fragment sequenced to bridge the gap between
            cosmids ZK643 and R08D7.
            Bases 1 to 86 of this entry correspond to bases 39449 to 39534 of
            cosmid ZK643 (EMBL:CEZK643, accession number Z11126). Bases 1758 to
            1762 of this entry correspond to bases 1 to 5 of cosmid R08D7
            (EMBL:CER08D7).

            NCBI gi: 6966
FEATURES             Location/Qualifiers
     source          1..1762
                     /organism="Caenorhabditis elegans"
BASE COUNT      487 a    403 c    459 g    413 t
ORIGIN
```

```
Initial Score    =    167  Optimized Score  =    403  Significance = 10.77
Residue Identity =    51%  Matches          =    493  Mismatches   =   328
Gaps             =    129  Conservative Substitutions               =     0
```

```
            X         10         20         30         40         50
```



```
          GTCGACTG--GAACGAGAC-GACCTGCTCCAGACCATTCAGCA--ATC--CTGC-TCTG--G
           |||  | |  | | || ||  | | |     |||   ||| ||  |||  || | ||||  |
ACTCGTCTTCGTCATCAGGCGGAGGATACTCATCGGGAGGTGACAGCTCATCATCATCATCTTCATCTGGAG
       200        210        220        230        240        250        260

          60         70         80         90        100        110
GCT---CCCAAGGGCTGAGATTACAGGTGTAACCGAGAT-AAGACCCT-GGCTG--AGTCAACGATGCACAA
| |   |  ||||  | ||||  ||   ||   ||||  | |  | ||| | |||   |  |   || | |
GATATTCGGGAGGG-AGTGATT-CA-TCGT---CGAGCTCATCATCCTCGTCTGGTGGATATTCATCCGGAG
      270        280            290        300        310        320

120        130            140        150        160        170        180
GGCTCGAATCCTAG-GGAT----GGGAGGGGAGGGGCACCCCAG--GGTGGAGCTCCCCCCTCCTTCTTGCC
|   || || || | | ||     |||||  || |  | |  | |  || |||  | ||   | ||||  |  |
GAGGCG-ATGCTGGAGCATCATCTGGAGGAGA-GTCCTCTTCTGCAGGAGGATATTCCGGAT-CTTCAT--C
330        340        350        360        370        380        390

     190        200        210        220        230        240        250
CTTTGCCAGCGAAG--GCCCCAGGCGGCTGC-CAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGC
    |  ||||   |  | || || | | |  | | |   || |||| |||||  |||| ||| |||
AAGCGGAGGAGAAGCATCATCCGGAGGATACTCTGGTGGATC---GTCAGAGCCAGCACCAGCTCCAGAAGC
  400        410        420        430            440        450        460

       260            270        280        290        300        310
AACAGCTGC-----CTGGGCCGAGGCTCTGGGAGCTCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCA
| || ||||      |||||  ||  ||| ||  | |||| ||  |   |  | |   ||||   |  |||
AGCACCTGCTTCATCTGGG-GGATACTCAGGAGGATCCGAAGCAGCACCAG--AAGCTGCTCC--AGCAGCA
   470        480        490        500        510            520        530

320        330        340        350        360        370        380
GCA--GGGTGGAGGAGTCCGG--CTGGCACTGCACCTCCAG--GCACCGAGAG-AAATTCTGGCAGGCCTTC
 ||   ||||||  | || ||  |||  |||||||   ||  || ||   ||      || ||  |  |||
CCATCAGGTGGA-TATTCAGGATCTGAGGCTGCACCAGAAGCTGCTCCAGCAGCCCCATCAGGTGGATATTC
       540        550        560        570        580        590        600

     390        400        410        420        430        440        450
CCGATACAGGGCTTCAGAGCAAGCAGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGG-TCTGGAC
  |||  |   |||  | || |||||  | |||  |  |  ||  |||| ||| ||   || ||||  |  |
AGGATCC---GAAGCAGCACCAGAAGCTG-CTCCAGCAG--CACCATCAGGT-GGATATTCAGGTTCTGAA-
         610        620        630            640        650        660

   460        470        480            490        500        510        520
GAATCGCAGACATTCTGGTAGGGGCT--GGAAGGTACATGAGGT-GACAAAATGTATCTCGGTGTTGACGTC
     || | ||  | || ||  |||   |  |   |||  ||| || |   || |||| |  |  |
-----GCTG-CA--CCGGAAGCCGCTCCAGCGGCCCCATCCGGTGGATACTCTGGATCT-GAAGCAG-----
       670            680        690        700        710        720

     530            540        550        560            570        580
CTCCAGAAGC-GTT--TGCATCTTAGACCCTGCCACAGTC---TTCAGTTGCTCTATCCAG-CGCTGGGCTA
| |||||||| | |   |||  ||     |   |  | | ||    ||||  ||    |||||  || | | |
CACCAGAAGCAGCTCCAGCAGCTCCATCTGGGGGATATTCCGGATCAGAGGC-AGCTCCAGAAGCCGCACCA
   730        740        750        760        770            780        790

590        600        610        620        630        640        650
GGAAGAGGCTCCA-CAAGGCCTTGCAGTGCTTCT-CGTCCTGAAGA-TTGACG-GCCACAGTGA-CTGGGTA
| |    || ||| |  ||   | |    | |||   |||| | |  || ||  ||| ||  | ||  ||
GCA----GCCCCATCTGGGGGATACTCAGGATCTGAATCCTCAGCACCTGCCGCTCCAGAGCCAGCTCCGTC
      800        810        820        830        840        850        860

  660        670        680        690        700        710        720
```

```
    ATCTTTAAGCAGGTGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGG-AGATGGGACTGTG--GCT
     |||  | ||   |    ||  |       |  ||  | |  |  | |||||| ||     ||| ||  | |
    TTCTGGAGGCTATTCTGGAG-GAGGAGGTGACGCTGGA-TCAGCTGCTGGAGGCAGCAACTACTCTGGAGGT
        870        880        890        900        910        920        930

       730        740        750        760        770        780        790
    GAAGTAACAG-TC-AGGTGTCCC-CCGC-AGGCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGGAAT
    |   | ||| | || || || ||| || | ||    ||| |||| | |     ||| | ||||  |  |
    GGTGAAACCGCTCCAGCTGCCCCACCACCAGCTCCGGAACCAGCTCAAACCTACTCAGGAGCAGGTGGAGAG
        940        950        960        970        980        990       1000

         800        810        820        830        840        850        860
    TTGGGCTCCAGGCTGGCGCCAGCACTGTCATCTCTGCCGGGGACCCCTCATGCCTGTGACAGAGCAGGTCG-
    | || | ||  ||  || ||||||||  |||||  |  | |||  |    ||  |||   |||| ||| |
    TCGGCCGCC--GC-CGCACCAGCAC---CATCT-GGAGGAGGATAC--AGTGGAAGTGGAGGAGCCGGTGGA
     1010       1020       1030       1040       1050       1060

             870          X
    TCTC------GTTCCAGT--CGAC
     | |       |  ||| |   |||
    GCACCAAATGGAGCCAATTATGACGAAGCACAAG
  1070      1080      1090 X    1100


6. US-08-162-407-1' (1-879)
   DMU13397     Drosophila melanogaster Oregon R DTIS11 protein mR

 LOCUS       DMU13397     2091 bp    mRNA            INV       30-AUG-1994
 DEFINITION  Drosophila melanogaster Oregon R DTIS11 protein mRNA, complete cds.
 ACCESSION   U13397
 KEYWORDS    .
 SOURCE      fruit fly.
   ORGANISM  Drosophila melanogaster
             Eukaryotae; Hyperchondria; Eukaryote crown group; Metazoa/Eumycota
             group; Metazoa; Eumetazoa; Bilateria; Coelomata; Protostomia;
             Arthropoda; Tracheata; Insecta; Pterygota; Neoptera; Holometabola;
             Diptera; Brachycera; Cyclorrhapha; Drosophilidae; Drosophila;
             Sophophora; melanogaster group; melanogaster subgroup.
 REFERENCE   1  (bases 1 to 2091)
   AUTHORS   Ma,Q., Wadleigh,D., Chi,T. and Herschman,H.
   TITLE     The Drosophila TIS11 homologue is a developmentally controlled gene
   JOURNAL   Oncogene (1994) In press
   STANDARD  full automatic
 REFERENCE   2  (bases 1 to 2091)
   AUTHORS   Ma,Q.
   TITLE     Direct Submission
   JOURNAL   Submitted (12-AUG-1994) Qiufu Ma, Molecular Biology, Bristol-Myers
             Squibb Pharmaceutical Research Institute, Road 206/Province line
             Road, Princeton, NJ 08543, USA
   STANDARD  full automatic
 COMMENT     NCBI gi: 532795
 FEATURES             Location/Qualifiers
      source          1..2091
                      /clone_lib="Drosophila eye disc lambda gt10 cDNA library"
                      /chromosome="X"
                      /strain="Oregon R"
                      /organism="Drosophila melanogaster"
                      /map="12B"
                      /dev_stage="instar, larval"
      CDS             301..1614
```



```
                      /note="NCBI gi: 532796"
                      /codon_start=1
                      /product="DTIS11 protein"
                      /translation="MSADILQKSREQDDSHYFERGDISKYVTMNDHLGDFDCNEVRKE
                      IRMLLAHGANLDQQHQQQPHRHHGGLTRTISQPAQLIQQQQQQHQQQQQQQQPPVASL
                      VTITENLGNMNLHRKLERTQSEPLPPQQPMNTSRYKTELCRPFEEAGECKYGEKCQFA
                      HGSHELRNVHRHPKYKTEYCRTFHSVGFCPYGPRCHFVHNADEARAQQAAQAAKSSTQ
                      SQSQSQQSSSQNFSPKSNQSSNQSSNSSSSSSSSGGGGGGGNSINNNNGSQFYLPLSP
                      PLSMSTGSDRESPTGSLSLSPTNSLTSFPFHDALQHGYLASNGAKSNSSASSTSSASG
                      MGLGMSMGIGQGMIIGQGLGMGHHGPATPPESPNVPISPVHTPPPYDVVVSGSGAGNN
                      SVGSKQLLQKSVSTPMQQEDTPRLPVFNRLSSGVEAYQQQSNLGL"
      misc_feature    724..897
                      /note="encodes cysteine/histidine repeats"
 BASE COUNT      658 a    580 c    472 g    381 t
 ORIGIN

Initial Score    =    145  Optimized Score  =    406  Significance =  9.04
Residue Identity =    52%  Matches          =    504  Mismatches   =   315
Gaps             =    139  Conservative Substitutions               =     0

             X         10         20         30         40
             GTCGACTG-GAACGA-GA--CGACCTGC-TCCAG----ACCATTCAGC--AATCC--TGCTC
               || |  | | | || ||  ||  || | || ||     | ||  | |  ||| |    |
    TTCTGCAGAAATCAAGAGAGCAGGATGATTCGCACTACTTCGAGCGTGGCGATATATCCAAATACGTAACGA
        320   X   330        340        350        360        370        380

   50         60         70         80         90        100
    TG--GGCTCCCAAGGGCTGA----GATTAC----AGGTG-TAACCGAGATAA-GACCCTGGCTGAGTCAACG
    ||   | || |  ||| |||    |||| |    |||||    |  || |||| ||  || |||  | | |||
    TGAACGATCACTTGGG-TGATTTCGATTGCAACGAGGTGCGCAAGGAAATAAGGATGCT-GCT-CGCCCACG
      390        400        410        420        430        440        450

 110        120        130        140        150        160        170        180
   ATGCACAAGGCTCGAATCCTAGGGATGGGAGGGGAGGGGCA-CCCCAGGGTGGAGCTCCCCCCTCCTTCTTG
     || |||  || | |||     | |   |   | ||   || | |||   ||| | |     |||   |
    GCGC-CAA--CTTGGATCAGCAGCACCAG-CAGCAGCCACATCGCCACCATGGCGGT-----CTCACACGCA
       460        470        480        490        500              510

          190        200        210        220        230        240        250
    CCCTTTGCCAGCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCA
    || |||  ||||   || ||||| |   | ||||||||| |||    |    || ||||||||||| ||||
    CCATTTCACAGC--CGG-CCCAG-CTCAT-CCAGCAGCAGCAGCAGCAACACCAACAGCAGCAGCAGCAGCA
    520          530          540        550        560        570        580

         260        270        280        290        300        310        320
    ACAGCTGCCTGGGCCGAGGCTCTGG-GAGC-TCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAG
     ||||  ||||   | ||  ||||| || | ||   |    |  |||| |   ||  |||   |  | | |
    GCAGCCACCTGTTGCCAG--TCTGGTGACCATCACCGAGAACCTGGGCAACATGA-ACCT---GCACCGAAA
       590        600        610        620        630        640        650

          330          340        350        360        370        380
    GGTGGAG-GAGTCC---GGCTGGCACTGCACCTCCAGGCACCGA-GA-GAAATTCTG--GCAGGCCTTCCCG
    | ||||| ||  ||    |  | |||||| |  |||   |||| ||  | || | |   || | |    |
    GCTGGAGCGAACCCAATCGGAGCCACTGCCGCCACAGCAGCCGATGAACACATCCAGATACAAGACCGAGCT
          660        670        680        690        700        710        720

     390        400        410        420        430        440        450
    ATACAGGGCTTCAGAGCAAGCAGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGGTC--TGGACGA
     | | |  | |  ||| | || |  |  ||||   || | |  || ||| |  ||| | |  ||  ||
    GTGCCGTCCGTTCGAGGAGGCCGGAGAATGCA--AGTACGGCGAGAAGTGCCA---GTTCGCCCATG---GA
```

```
       730       740       750       760       770       780

     460       470       480       490       500       510       520
ATCGCAGACATTCTGGTAGGGGCTGGAAGGTACAT--GAGGTGACAAAATGTATCTCGGTGTTGACGTCCTC
| |   ||  ||  |  | |   || |  || |||   | || |||| | | |   |  ||  | |  | ||
AGCCATGA-GTTGCGAAACG---TGCACCGT-CATCCCAAGT-ACAAGACGGAATAC--TGCCG-CACCTTC
 790       800       810       820       830       840

 530       540       550       560       570       580       590
CAGAAGCGTTTGCATCTTAGACCCTGC-CACAG-TCT-TCAGTTGCTCTATCCAGCGCTGGGCTAGGAAGAG
||  |||||  || |||  | |||| |  || |  || ||| ||  | |   ||  || || | |||
CA-CAGCGTGGGCTTCT--GTCCCTACGGACCGCGCTGTCACTT--TGTTCACAATGC-GGACGAGGCCCGC
850       860       870       880       890       900       910

   600       610       620       630       640       650       660
GCTCCACAAGGCCTTGCAGTGCTTCTCGTCCTGAA--GATTGACGGCCACAGTGACTG-GGTAATCTTTAAG
|| |||  ||||   |||     |  |||||  |   | |  | | | ||||  | |  ||  || |   |
GC-CCAACAGGCGGCCCAGGCAGCCAAGTCCTCCACCCAGTCGCAGTCGCAGTCGCAGCAGTCGTCGT--CG
   920       930       940       950       960       970       980

   670           680       690       700       710       720       730
CAGGTGGTC-AGTCAA----CTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGACTGTGGCTGAAGTAA
|||    ||  | | |    |   |  | |  ||  |  |||| | |  || | | |  |  || | || |
CAGAACTTCTCGCCGAAGAGCAACCAGAGCAGCA-ATCAAAGTAGCAACAG-TAGCA--GCAGCAGCAGCAG
   990      1000      1010      1020      1030          1040      1050

       740       750       760       770       780       790
CAGTCAGGTGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGGAATT-----TG-GG
|||   || |  |  ||  |||  ||   || ||||| ||||| || |||| |    ||||     ||  |
CAG--CGGCG-GCGGCGGTGGC--GGCGGCAACAGCATCAACAACAACAACGGTAGCCAATTCTATCTGCCG
      1060      1070      1080      1090      1100      1110

    800       810       820       830           840           850
CT----CCA-GGCTG-GCGCCAGCACTGTCATCTCTGCCGGGGA---CCC-C----TCA-TG--CCTGTGAC
||    |||  |||| ||   ||||| |  |||    ||||| |   ||| |    ||| ||  |||  | |
CTAAGCCCACCGCTGAGCATGAGCACAG-GATC-GGACCGGGAATCGCCCACCGGATCACTGTCCCTCAGCC
1120      1130      1140      1150      1160      1170      1180

       860           870         X
AGAGCAGGTCGT---CTCG-TTCCAG-TCGAC
  | ||  ||||   |  | |||| | || ||
CCACCAACTCGTTGACCAGCTTCCCGTTCCACGATGCCCTGC
1190      1200      1210      1220      1230
```

7. US-08-162-407-1' (1-879)
   CRUPA1V Chinese hamster pro-alpha-1 (V) collagen mRNA, com

LOCUS CRUPA1V 6114 bp ss-mRNA ROD 14-JAN-1992
DEFINITION Chinese hamster pro-alpha-1 (V) collagen mRNA, complete cds.
ACCESSION M76730
KEYWORDS alpha-1 type V collagen.
SOURCE Cricetulus longicaudatus male adult cDNA to mRNA.
ORGANISM Cricetulus longicaudatus
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Rodentia; Myomorpha; Cricetidae; Cricetinae; Cricetini.
REFERENCE 1 (bases 1 to 6114)
AUTHORS Greenspan,D.S., Cheng,W. and Hoffman,G.G.
TITLE The pro-alpha-1(V) collagen chain: Complete primary structure,
distribution of expression, and comparison with the pro-alpha-1(XI)

collagen chain
JOURNAL J. Biol. Chem. 266, 24727-24733 (1991)
STANDARD full automatic
COMMENT NCBI gi: 191150
FEATURES Location/Qualifiers
source 1..6114
/organism="Cricetulus longicaudatus"
/dev_stage="adult"
/sequenced_mol="cDNA to mRNA"
/sex="male"
CDS 33..5555
/gene="pro-alpha-1 (V) collagen"
/note="NCBI gi: 191151"
/codon_start=1
/product="pro-alpha-1 type V collagen"
/translation="MDVHTRWKDRLPVGPAAVPPLLLLLLLLWAPPQSRAAQPTDLLE
MLDFHNLPSGVTKTTGFCATRRSSRGPDVAYRVSKDAQLSMPRKQLYPDSDFAEDFSI
LTTVKAKKGSQAFLVSVYNEQGIQQVGMELGRSPVFLYEDHTGKPGPEEYPLFPGINL
SDGKWHRIAISVYRKNVTLILDCKKKVVKFLNRSDHPIIDVNGIIMFGSRILDDEIFE
GDIQQLLFVSDHRAAYDYCEHYSPDCDTAVPDTPQSQDPNPDEYYPEGDGETYYYEYP
YYEDPEDLGKEPAPTQKPVEAARETTEVPEEQTQPPPEAPTVPETSDPAGKEDDPGFG
DYDYVPTDDYYTPAPYEDLGYGEGVENPDQPTNPDSGAEVPTSTIITSNTSNPAPSPG
EDKDDLGGEFTEETIKNLEENYYDPYFDPDSDSNVSPSEIGPGMPANQDTIYEGIGGP
RGEKGQKGEPAIIEPGMLIEGPPGPEGPAGLPGPPGTTGPTGQVGDPGERGPPGRPGL
PGADGLPGPPGTMLMLPFRFGGGGDAGSKGPMVSAQESQAQAILQQARLALRGPAGPM
GLTGRPGPMGPPGSGGLKGEPGDMGPQGPRGVQGPPGPTGKPGRRGRAGSDGARGMPG
QTGPKGDRGFDGLAGLPGEKGHRGDPGPSGPPGLPGDDGERGDDGEVGPRGLPGEPGP
RGLLGPKGPPGPPGPPGPVGTGMDGQPGPKGNVGPQGEPGPPGQQGNPGAQGLPGPQGAI
GPPGEKGPLGKPGLPGMPGADGPPGHPGKEGPPGEKGGQGPPGPQGPIGYPGPRGVKG
ADGIRGLKGTKGEKGEDGFPGFKGDMGIKGDRGEIGPPGPRGEDGPEGPKGRGGPNGD
PGPLGPTGEKGKLGVPGLPGYPGRQGPKGSIGFPGFPGANGEKGGRGTPGKPGPRGQR
GPTGPRGERGPRGITGKPGPKGNSGGDGPAGPPGERGPNGPQGPTGFPGPKGPPGPPG
KDGLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGERGHPGPPGPPGEQGL
PGVAGKEGTKGDPGPAGLPGKDGPPGLRGFPGDRGLPGPVGALGLKGSEGPPGPPGPA
GSPGERGPAGAAGPIGIPGRPGPQGPPGPAGEKGVPGEKGPQGPAGRDGLQGPVGLPG
PAGPVGPPGEDGDKGEIGEPGQKGSKGDKGEQGPPGPTGPQGPIGQPGPSGADGEPGP
RGQQGLFGQKGDEGSRGFPGPPGPVGLQGLPGPPGEKGETGDVGQMGPPGPPGPRGPS
GAPGADGPQGPPGGIGNPGAVGEKGEPGEAGEPGLPGEGGPLGPKGERGEKGEVGPSG
AAGPPGPKGPPGDDGPKGSPGPVGFPGDPGPPGEPGPAGQDGPPGDKGDDGEPGQTGS
PGPTGEPGPSGPPGKRGPPGPAGPEGRQGEKGAKGEAGLEGPPGKTGPIGPQGAPGKP
GPDGLRGIPGPVGEQGLPGSPGPDGPPGPMGPPGLPGLKGDSGPKGEKGHPGLIGLIG
PPGEQGEKGDRGLPGPQGSSGPKGEQGITGPSGPLGPPGPPGLPGPPGPKGAKGSSGP
TGPKGEAGHPGLPGPPGPPGPPGEVIQPLPIQASRTRRNIDASQLLDDGAGESYLDYADGM
EEIFGSLNSLKLEIEQMKRPLGTQQNPARTCKDLQLCHPDFPDGEYWVDPNQGCSRDS
FKVYCNFTAGGSTCVFPDKKSEGARITSWPKENPGSWFSEFKRGKLLSYVDAEGNPVG
VVQMTFLRLLSASAHQNITYNCYQSVAWQDAATGSYDKAIRFLGSNDEEMSYDNNPYI
RALVDGCATKKGYQKTVLEIDTPKVEQVPIVDIMFNDFGEASQKFGFEVGPACFLG"
BASE COUNT 1372 a 1768 c 1883 g 1091 t
ORIGIN

Initial Score = 144 Optimized Score = 408 Significance = 8.96
Residue Identity = 52% Matches = 508 Mismatches = 301
Gaps = 157 Conservative Substitutions = 0

```
          X       10        20        30        40        50
          GTCGACTGGAACGAGACGACCTGCTCCAGACCATTCA-GCAATCCTGCT---CTGGG-CTCC
          | | | ||| | ||| ||       || | || | || ||   |||| |   | ||| || |
GGAGAAACCGGCCCAATGGGA-GAG-CGTGGCCACCCTGGCCCTCCAGGCCCCCCTGGTGAACAGGGCCTTC
  3060  X 3070        3080      3090      3100      3110      3120
```

```
  60        70        80        90        100       110       120
CAAGGGCTGAGATTACAGGTGTAACCGAGATAAGACCCTGG--CTG-AGTCAACGATGCACAAGGCTCGAAT
CAGGAGTTGCTGGAAAAGAAGGGACTAAG-GGGGACCCAGGCCCTGCAGGCCTCCCTGGA-AAGGATGGCCC
   3130      3140      3150      3160      3170      3180      3190

 130       140       150       160       170       180
CCTAGGGATG---GGAGGGGAGGGGCACCCCAGGGTGGAGCTCCC--CCCT-----CCTTCTTGCCCT--TT
TCCAGGATTGCGTGGATTCCCTGGGGACC-----GAGGACTTCCCGGCCCTGTGGGAGCTCTTGGGCTAAAA
   3200      3210      3220      3230      3240      3250      3260

 190       200       210       220       230       240       250
GCCAGCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCAACAGCT
GGCAGCGAAGGCCCCCCG-GGC--CCTCCAGGTCCTGCGGGTTCTCCAGGAG-AG-AG-AGGACCAGCTGGG
      3270      3280      3290      3300      3310      3320

260       270       280       290       300       310       320
GC--CTGGGCCGAGGCTCTGGGAGCTCC-GTGGCTTCTAGGGCTATGGGA-CTCCTTGGGGGCAGCAGG-GT
GCTGCTGGGCCCA---TC--GGAATTCCAGGGAGACCTGGGCCTCAGGGACCTCC--AGGGCCAGCTGGAGA
3330      3340      3350      3360      3370      3380      3390

   330       340       350       360       370       380       390
G-GAGGAGTCCGGCTGGC---ACTGCACCTCCAGGCACCGAGAGAAATTCTGGCAGGCCTTCCCGATAC---
GAAAGGGGTTC--CTGGCGAGAAAGGTCCACAAGGCCC-----AGCTGGCCGAGATGGCCTCCAAGGTCCTGT
     3400      3410      3420      3430      3440      3450

       400       410       420       430       440       450       460
AGGGCTTCAGAGCAAGCAGCTGTGTGCAGGT-GTCCTTCAGGAGGTGGGAGATGTTGGTCTGGACGA-ATCG
GGGGCTTCCTGG--TCCAGCCG-GCCCCGTTGGTCCTCCTGGA----GAAGATGGAGATAAGGGAGAGATCG
3460      3470      3480      3490      3500      3510      3520

       470       480       490       500       510       520
CAGACATTCTGGTAGGGGCTGGAAGGTA-CATGAGGTGACAAAATGTATCTCGGT----GTTGAC------G
GGGA------GCCA-GGGCAGAAAGGAAGCAAG-GGGGACAAAGGCGAGCAGGGTCCTCCTGGACCTACGGG
          3530      3540      3550      3560      3570      3580

       530       540       550       560       570       580
TCCTCCAGAAGCGTTTG--CATCTTAGACCCTGCCACAGTCTTCAGTTGCTCTATCCAG---CGCTGGGCTA
TCCTCAAGGCCCCATTGGACAGCCAGGTCCCT-CGGGAG----CAGATGGT-GAGCCAGGTCCTCGTGGACA
3590      3600      3610      3620      3630      3640      3650

590       600       610       620       630       640       650
GGAAGAGGCTCCACAAGGCCTTGCAGTGCTTCTCGTCCTGAAGA--TT---GACGGCCACAGTGACTGGGTA
GCAAG-GCCTGTTC-GGGC--AGAAAGGAGACGAAGGCTCAAGAGGTTTCCCAGGACCACCGGGGC-CAGTG
     3660      3670      3680      3690      3700      3710

  660       670       680       690       700       710       720
ATCTTTAAGCAGGTGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGACTGTGGCTGAA
GGATTACAG-GGTTTGCCAG-GACCTC-CAGGAGAGAAGGGTGA---GACAGGAGATG----TGGGCCAGAT
3720      3730      3740      3750      3760      3770      3780
```



```
  730       740       750       760       770       780
GTAACAGTCAGGTGTCCCCC--GCAGGCAAGGA--CTCAGCAGCAGCAACAGC-----AGCAACAG------
GGGAC-CTCCGG-GTCCCCCAGGCCCCCGAGGACCCTCTGGAGCTCCAGGTGCTGATGGGCCACAGGGTCCT
     3790      3800      3810      3820      3830      3840      3850

     790       800       810       820       830       840
---GGAGGAATTTGGGCTCCAG--GCTG-GCGCCAGCACTGTCATCTCT------GCCGGGGACCC--CTCA
CCCGGAGGAATTGGCAACCCTGGTGCAGTGGGAGAAAAGGGGGAGC-CTGGTGAAGCTGGAGAGCCTGGTC-
     3860      3870      3880      3890      3900      3910      3920

     850       860       870       X
TGCCTGTGACAGAGCAGGTCGTCTCGTTCC---AGTCGAC
TTCCAG-GAGA-AGGAGGTCCCCTGGGACCTAAAGGAGAAAGAGGGGAGA
     3930      3940      3950      3960
```

8. US-08-162-407-1' (1-879)
   GORINVOLUB Gorilla gorilla involucrin gene medium allele, com

```
LOCUS       GORINVOLUB   1818 bp ds-DNA             PRI       08-SEP-1994
DEFINITION  Gorilla gorilla involucrin gene medium allele, complete cds.
ACCESSION   M23604 J04499
KEYWORDS    epidermal protein; involucrin.
SOURCE      Gorilla gorilla (individual_isolate Gorilla K) (library: vagina)
            DNA.
  ORGANISM  Gorilla gorilla
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Pongidae.
REFERENCE   1  (bases 1 to 1818)
  AUTHORS   Teumer,J. and Green,H.
  TITLE     Divergent evolution of part of the involucrin gene in the
            hominoids: Unique intragenic duplications in the gorilla and human
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86, 1283-1286 (1989)
  STANDARD  full automatic
COMMENT     Draft entry and computer readable copy of sequence [1] kindly
            provided by J.K. Teumer (03-31-89).

            NCBI gi: 340978
FEATURES             Location/Qualifiers
     source          1..1818
                     /organism="Gorilla gorilla"
                     /isolate="Gorilla K"
                     /cell_type="keratinocyte"
                     /sequenced_mol="DNA"
                     /clone="pIGorH6.1"
                     /tissue_lib="vagina"
     CDS             1..1818
                     /note="NCBI gi: 536828"
                     /codon_start=1
                     /product="involucrin"
                     /translation="MSLQHTLPVTLSPALSQELLKTVPPPVNTQQEQMKLPTPLPPPC
                     QKVPVKLPVEVPSKQEEKHMTAVKGLPEQECEQQQQEPQEQEPQQQHWEQHEEHQKAE
                     NPEQQLKQEKAQRDQQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELPEQRE
                     GQLKHLEQQKGQLELPEQQEGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLELPEQ
                     QEGQLELPEQQEGQLKHLEQQEGQLELPEQQEGQLELPEQQEGQLKHLEQQEGQLKHL
                     ENQEGQLEVPEEQVGQLKYLEQQEGQLKHLDQQEKQPELPEQQVGQLKHLEQQEGQPK
                     HLEQQEGQLEHLEEQEGQLKHLEQQEGQLEHLEHQEGQLGLPEQQVRQLKQLEKEEGQ
```

```
                     PKHLEEEEGQLKHLVQQEGHVEHLVQQEGQLEHLVQQEGQLEQQEGQVEHLEQQVGQL
                     KHLEEQEGQLKHLEQQQGQLGVPEQQVGQPNNLEQEEKQLELPEQQEGQLKHLKKQEA
                     QLELPEQQVGQPKHLEQQEKQLEHPEQQDGQLKHLEQQEGQLKDLEQQKGQLEQPVFA
                     PAPGQVQDIQPALPTKGELLLPIEHQQQKQEVQWPPKHK"
BASE COUNT      531 a    454 c    657 g    176 t
ORIGIN

Initial Score    =    141  Optimized Score  =    407  Significance =  8.73
Residue Identity =    52%  Matches          =    499  Mismatches   =   329
Gaps             =    120  Conservative Substitutions               =     0

             X       10          20          30          40          50
             GTCGACTGGA-ACGAGACG-ACCT---GCTCCAGACCATTCAGCAATCCTGCT--CTGGGCT
             |  |   ||| |   || | ||||   ||  ||||      ||||  |   |||  | | ||
   TCCCAGAGCAGCAGGAGGGACAGCTGAAGCACCTGGAGCAGCAGAAGGGGCAGC--TGGAGCTCCCAGAGCA
       560   X   570         580         590         600         610         620

        60          70          80          90          100         110
   CC--CAAGGGC----TGAGATTACAGGTGTAACCGAGATAAGACCCTG--GCT----GAGTCA--ACGATGC
     |   | ||||     ||| |  || | | | | ||    |   |||  |||     ||| ||  | ||| |
   GCAGGAGGGGCAGCTGGAGCTCCCA-GAGCAGCAG-GAGGGGCAGCTGGAGCTCCCAGAG-CAGCAGGAGGG
       630         640         650         660         670         680         690

       120         130         140         150         160         170         180
   ACAAGGCTCGAA--TCCTAGGGATG--GGAGGGG-AGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTG
   |||  ||| |||   |||||  |   |  |||||||  ||   | | |   |  | |||                |  || |
   ACA--GCT-GAAGCACCTAGAGCAGCAGGAGGGGCAGCTGGAGCTCCCAGAGCAGC-AGGAGGGGCAGCTGG
           700         710         720         730         740         750         760

           190         200         210         220         230         240
   CCCTTTGCCAGCGAAGGCCCCAGG--CGGCTG--CCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGG
     |  | ||||  | | ||   |||  | ||||   || | | | |||   ||| ||||||  | ||||  | ||
   AGC--TCCCAGAGCA-GCAGGAGGGACAGCTGAAGCACCTGGAGCAGCAGGAGGGGCAGCTGAAGCATCTGG
             770         780         790         800         810         820

    250         260         270         280         290         300         310
   AGCAACAGCTGCCTGGGC-CGAGG-CTCTG-GGAGC-TCCGTGGCTTCT-AGGGCTATGGGACTCCTTG-GG
   ||  | |||   |      ||  ||||  | |  | |||||     | |||  || | |    ||| ||  |   | ||
   AGAACCAGGAGGGGCAGCTGGAGGTCCCAGAGGAGCAGGTGGGGCAGCTGAAGTACCTGGAACAGCAGGAGG
  830         840         850         860         870         880         890         900

       320         330         340         350         360         370         380
   GGCAGC--AGGGTGGAGGAGTCCGGCTGGCACTGCACCTCCAGGCACCGAGAGAAATTCTGGCAGGCCTTCC
   ||||||  | |     ||| ||  ||  ||  |||  | |   || ||  |||| |   | ||  || |    |
   GGCAGCTGAAGCACCTGGA-TC-AGCAGGAGAAGCAGC-CGGAGCTCCCAGAGCA--GCAGGTGGGGC-AGC
         910         920         930         940         950         960

       390         400         410         420         430         440         450
   CGATACA---GG-GCTTCAGAGCAAGCAGCTGTGTGCAGGT-GTCCTTCAGGAGGTGGGAGATGTTG-GTCT
     ||  ||   || ||  |||      |||||  |  |||  | |  |  |||||||  || |  ||  |   ||
   TGAAGCACCTGGAGCAGCAGGAGGGGCAGC-CTAAGCATCTGGAGCAGCAGGAGG-GGCAACTGGAGCACCT
   970         980         990         1000        1010        1020        1030

          460         470         480         490         500         510         520
   GGACGAATCGCA--GACATTCTGGTAGGGGCTGGAAGGTACATGAGGTGACAAAATGTATCTCGGTGTTGAC
   |||  ||    |  |  ||   || | ||    |||| ||    || |||| | ||    || | | |   ||  ||
   GGAGGAGCAGGAGGGGCA-GCT-GAAGCACCTGG-AGCAGCAGGAGG-GGCA-GCTGGAGCAC-CTG--GA-
  1040        1050        1060        1070        1080        1090        1100

          530         540         550         560         570         580
```



```
   GTCCTCCAGAAGCGTTTGCATCTTAGACCCTGCCACAGTCTTCAGTTGCTCTA--TCCAGC-GCTGGGCTAG
   |  | ||||  |  |   ||| ||  |   || ||| || |  ||| |||    |  |  ||| ||| |   ||
   G--CACCAGGAAGG---GCAGCT--GGGGCTCCCAGAG-CAGCAGGTGCGGCAGCTGAAGCAGCTAGAGAAG
           1110        1120        1130        1140        1150        1160

 590         600         610         620         630         640         650
   GAAGAGGCTC-CACAAGGC-CTTGCA-GTGCTTCTCGTCCTGAAGATTGAC--GG-CCA-CAGTGACTGGGT
   || ||||   |   ||| || | || | | |       |  | |  ||   ||  ||  || ||| ||  |||
   GAGGAGGGGCAGCCAAAGCACCTGGAGGAGGAGGAGGGGCAGTTGAAGCACCTGGTGCAGCAG-GA-GGGGC
     1170        1180        1190        1200        1210        1220        1230

        660         670         680         690         700         710         720
   AATCTTTAAGCAGGTGGT-CAGTCAACTCTCTAAACTTCA-CTT-TGAAGTTGGAGGAGATGGGACTGTGGC
   |  | |  ||||  |||| ||| ||         | ||  | | | ||  |  | |||||  |   |||  ||
   A--CGTGGAGCACCTGGTGCAG-CAGGAGGGGCAGCTGGAGCATCTGGTGCAGCAGGAGGGGCAGCTGGAGC
       1240        1250        1260        1270        1280        1290        1300

        730         740         750         760         770         780
   TGAAGTAACAGTCAGGTG--TCCCCCGCAG--GCAAG-GACTCAGC---AGCAGCAACAGCAGCAACAGGGA
    | ||  |  | |||||   | || | ||  ||| | |  ||||  |||| | | || ||||  ||||
   AGCAG-GAGGGGCAGGTGGAGCACCTGGAGCAGCAGGTGGGGCAGCTGAAGCACCTAGAGGAGCAGGAGGG-
       1310        1320        1330        1340        1350        1360        1370

   790         800         810         820         830         840         850
   GGAATTTGGGCTCCAGGCTGGCGCCAGCACTGTCATCTCTGCCGGGGACCC--CTCATGCCTGTGACAGAGC
   | || |   ||  | ||   ||  |||||  | ||  |      |||| |||    || ||  |||    |||
   GCAACTGAAGCATCTGG--AGCAGCAGCAGGGGCAGTT-----GGGGGTCCCAGAGCA-GCAGGTGGGGCAGC
      1380        1390        1400        1410        1420        1430

   860         870         X
   AGGTCGTCTCGTTCCAGTCGAC
      |  |  |   |||  |
   CAAACAACCTGGAGCAGGAGGAGAAGCAGCTG
 1440        1450        1460        1470
```

9. US-08-162-407-1' (1-879)
   HUMCA1V Human mRNA for collagen alpha 1(V) chain, complete

```
LOCUS       HUMCA1V      5676 bp ss-mRNA            PRI       29-AUG-1991
DEFINITION  Human mRNA for collagen alpha 1(V) chain, complete cds.
ACCESSION   D90279
KEYWORDS    alpha 1(V) chain; collagen.
SOURCE      Human placenta, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
            Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 5676)
  AUTHORS   Takahara,K., Sato,Y., Okazawa,K., Okamoto,N., Noda,A., Yaoi,Y. and
            Kato,I.
  TITLE     Complete primary structure of human collagen alpha1(V) chain
  JOURNAL   J. Biol. Chem. 266, 13124-13129 (1991)
  STANDARD  full automatic
COMMENT     These data kindly submitted in computer readable form by: Kazuhiko
            Takahara
            Takara Shuzo Co., Ltd.
            Biotechnology Research Laboratories
            3-4-1 Seta, Otsu
            Shiga 520-21
            Japan
```

```
            Phone: 81-775-43-7200
            Fax:   81-775-43-2494.

            NCBI gi: 219509
  FEATURES             Location/Qualifiers
       source          1..5676
                       /organism="Homo sapiens"
       sig_peptide     127..237
                       /note="signal peptide of collagen alpha 1(V) chain"
                       /codon_start=1
       CDS             127..5643
                       /note="collagen alpha 1;  (V) chain precursor;  NCBI gi:
                       219510"
                       /codon_start=1
                       /translation="MDVHTRWKARSALRPGAPLLPPLLLLLLWAPPPSRAAQPADLLK
                       VLDFHNLPDGITKTTGFCATRRSSKGPDVAYRVTKDAQLSAPTKQLYPASAFPEDFSI
                       LTTVKAKKGSQAFLVSIYNEQGIQQIGLELGRSPVFLYEDHTGKPGPEDYPLFRGINL
                       SDGKWHRIALSVHKKNVTLILDCKKKTTKFLDRSDHPMIDINGIIVFGTRILDEEVFE
                       GDIQQLLFVSDHRAAYDYCEHYSPDCDTAVPDTPQSQDPNPDEYYTEGDGEGETYYYE
                       YPYYEDPEDLGKEPTPSKKPVEAAKETTEVPEELTPTPTEAAPMPETSEGAGKEEDVG
                       IGDYDYVPSEDYYTPSPYDDLTYGEGEENPDQPTDPGAGAEIPTSTADTSNSSNPAPP
                       PGEGADDLEGEFTEETIRNLDENYYDPYYDPTSSPSEIGPGMPANQDTIYEGIGGPRG
                       EKGQKGEPAIIEPGMLIEGPPGPEGPAGLPGPPGTMGPTGQVGDPGERGPPGRPGLPG
                       ADGLPGPPGTMLMLPFRFGGGGDAGSKGPMVSAQESQAQAILQQARLALRGPAGPMGL
                       TGRPGPVGPPGSGGLKGEPGDVGPQGPRGVQGPPGPAGKPGRRGRAGSDGARGMPGQT
                       GPKGDRGFDGLAGLPGEKGHRGDPGPSGPPGPPGDDGERGDDGEVGPRGLPGKPGPRG
                       LLGPKGPPGPPGPPGVTGMDGQPGPKGNVGPQGEPGPPGQQGNPGAQGLPGPQGAIGP
                       PGEKGPLGKPGLPGMPGADGPPGHPGKEGPPGEKGGQGPPGPQGPIGYPGPRGVKGAD
                       GIRGLKGTKGEKGEDGFPGFKGDMGIKGDRGEIGPPGPRGEDGPEGPKGRGGPNGDPG
                       PLGPPGEKGKLGVPGLPGYPGRQGPKGSIGFPGFPGANGEKGGRGTPGKPGPRGQRGP
                       TGPRGERGPRGITGKPGPKGNSGGDGPAGPPGERGPNGPQGPTGFPGPKGPPGPPGKD
                       GLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGERGHPGPPGPPGEQGLPG
                       LAGKEGTKGDPGPAGLPGKDGPPGLRGFPGDRGLPGPVGALGLKGNEGPPGPPGPAGS
                       PGERGPAGAAGPIGIPGRPGPQGPPGPAGEKGAPGEKGPQGPAGRDGLQGPVGLPGPA
                       GPVGPPGEDGDKGEIGEPGQKGSKGDKGEQGPPGPTGPQGPIGQPGPSGADGEPGPRG
                       QQGLFGQKGDEGPRGFPGPPGPVGLQGLPGPPGEKGETGDVGQMGPPGPPGPRGPSGA
                       PGADGPQGPPGGIGNPGAVGEKGEPGEAGEPGPSGRSGPPGPKGERGEKGESGPSGAA
                       GPPGPKGPPGDDGPKGSPGPVGFPGDPGPPGEPGPAGQDGPPGDKGDDGEPGQTGSPG
                       PTGEPGPSGPPGKRGPPGPAGPEGRQGEKGAKGEAGLEGPPGKTGPIGPQGAPGKPGP
                       DGLRGIPGPVGEQGLPGSPGPDGPPGPMGPPGLPGLKGDSGPKGEKGHPGLIGLIGPP
                       GEQGEKGDRGLPGPQGSSGPKGEQGITGPSGPIGPPGPPGLPGPPGPKGAKGSSGPTG
                       PRGEAGHPGPPGPPGPPGEVIQPLPIQASRTRRNIDASQLLDDGNGENYVDYADGMEE
                       IFGSLNSLKLEIEQMKRPLGTQQNPARTCKDLQLCHPDFPDGEYWVDPNQGCSRDSFK
                       VYCNFTAGGSTCVFPDKKSEGARITSWPKENPGSWFSEFKRGKLLSYVDAEGNPVGVV
                       QMTFLRLLSASAHQNVTYHCYQSVAWQDAATGSYDKALRFLGSNDEEMSYDNNPYIRA
                       LVDGCATKKGYQKTVLEIDTPKVEQVPIVDIMFNDFGEASQKFGFEVGPACFMG"
       mat_peptide     238..5640
                       /note="mature peptide of collagen alpha 1(V) chain"
                       /codon_start=1
  BASE COUNT     1181 a   1803 c   1872 g    820 t
  ORIGIN

Initial Score     =    139  Optimized Score  =    397  Significance =  8.57
Residue Identity  =    50%  Matches          =    472  Mismatches   =   348
Gaps              =    109  Conservative Substitutions               =     0

             X          10         20         30         40         50
             GTCGACTG---GAACGAGACGA-CCTGCTCCAGACCATTCAGCAATC-CTGCTC-TGGGCTC
    GGAGAAACGGGCCCAATGGGTGAGCGTGGCCACCCTGGGCC---CCCTGGACCCCCCGGTGAACAGGGGCTT
```



```
      3150 X    3160       3170       3180          3190       3200       3210

  60          70         80         90         100        110        120
  CCAAGGGC--TGAGATTACAGGTGTAACCGAGATAAGACCCTGG--CTG-AGTCAACGATGCACAAGGCTCG
  CC--GGGCCTTGCTGGAAAAGAAGGGA-CGAAGGGTGACCCAGGCCCTGCAGGCCTCCCTG-GGAAAGATGG
        3220       3230       3240       3250       3260       3270

     130        140        150        160        170        180        190
  AATCCTAGGGATGGGAGG---GGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCC--TTCTTGCCCTTTGCCA
  CCCTCCAGGATTACGTGGTTTCCCTGGGGA-CCGAGG-----GCTTCCTGGTCCAGTGGGAGCTCTTGGACT
3280       3290       3300       3310           3320       3330       3340

        200        210        220        230        240        250        260
  GCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGGAGCAACAGCTGCCT
  G-AAAGGCAATGAAGGGCCCCCTGGCCCACCAGGCCCTGCGGGATCTCCAGGGGAGAGAGGTCCAGCTGGAG
     3350       3360       3370       3380       3390       3400       3410

      270        280        290        300        310        320        330
  GGGCCGAGGCTCTGGGAGCTCC-GTGGCTTCTAGGGCTATGGGACTCCTTGGG--GGCAGCAGGGTGGAGGA
  CCGCTGGGCCCATCGGAATTCCAGGGAGACCTGGGCCCCAGGGACCCCCAGGGCCGGCAGGAGAGAAAGGGG
  3420       3430       3440       3450       3460       3470       3480

        340        350        360           370        380        390
  GTCC-GGCTGGCACTGCACCTCCAGGCACCG------AGAGAAATTCTGGCAGGCCTTCCCGATACAGGGCTT
  CTCCTGGCGAGAAAGGC-CCACAAGGCCCAGCTGGCCGAGACGGTCT-CCAGG--GGCC---TGTGGGGCTC
3490       3500       3510       3520       3530           3540       3550

  400        410        420        430        440        450        460
  CAGAG-CAAGC-AGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGGTCTGGACGAATCGCAGACAT
  CCGGGTCCAGCTGGCCCTGTG--GGTCCCCCTGGAGAAGACGGAGATAAGGG---AGA-G-ATCGGGGA---
     3560       3570       3580       3590       3600           3610

  470        480        490        500        510        520        530
  TCTGGTAGGGGCTGGAAGGTA-CATGAGGTGACAAAATGTATCTCGGT-GTTGACGTCCTCCAGAAGC----
  ----GCCGGGGCAGAAAGGAAGCAAG-GGGGACAAAGGAGAACAGGGTCCTCCTGGGCCTACAGGTCCTCAA
        3620       3630       3640       3650       3660       3670       3680

     540        550        560        570        580        590        600
  GTTTGCATCTTAGACCCTGCCACAGTCTTCAGTTG--CTCTATCCAG---CGCTGGGCTAGGAAGAGGCTCC
  GGCCCCATCGGACAGCCAGGC-CCCTCTGGAGCTGACGGCGAGCCGGGGCCTCGGGGCCA-GCAGGGCCTTT
      3690       3700       3710       3720       3730       3740       3750

        610        620        630        640        650        660
  ACAAGGCCTTGCAGTGCTTCTCGT-CCTGAAGATTGAC-GGCCACAGTGACTGGGTAATCTTTAAGCAGG--
  TC-GGGCAGAAAGGTGATGAAGGTCCCAGAGGCTTTCCTGGACCCCCTG---GGCCAGTGGGGCTGCAGGGT
        3760       3770       3780       3790           3800       3810

670        680        690        700        710        720        730
  TGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGAC-TGTGGCTGAAGTAACAGTCAGG
  TTGCCAG--GAC-CTCCAGGC------GAGAAGGGTGA-GACAGGAGACGTGGGCCAGATGGGCCCCCCGGG
```

```
3820         3830            3840       3850       3860       3870       3880

 740       750       760       770       780       790       800
   TGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGGAATTTGGGCTCCAGGCTGG---
   |  |||  ||   | ||||| |  | | |||  |||  ||   | ||   |    |  | ||||| |||
   TCCCCCTGGCCCCCGAGGAC-CCTCCGGAGCTCCAGGTGC-TGATGGCCCACAAGGTCCCCCAGG-TGGAAT
        3890      3900       3910      3920       3930       3940       3950

  810       820       830       840       850       860          870
   CGCCAGCACTGTCATCTCTGCCGGGGACCCCTCATGCCTGTGACAGAGCAGGT-CGTCT---CGTTCCAGTC
    |  | | |||    |  ||   | ||     |  ||||| |  | |||||||  | ||   | |||| |
   AGGAAACCCTG-GTGCAGTGGGAGAGA-AGGGCGAGCCTG-GCGA-AGCAGGTGAGCCTGGCCCTTCCGGGA
         3960       3970       3980       3990       4000       4010

     X
   GAC
   ||
   GAAGCGGCCCCCC
 4020      4030
```

10. US-08-162-407-1' (1-879)
    HUMPA1V Human pro-alpha-1 (V) collagen mRNA, complete cds.

```
LOCUS       HUMPA1V      7138 bp ss-mRNA           PRI       07-JAN-1995
DEFINITION  Human pro-alpha-1 (V) collagen mRNA, complete cds.
ACCESSION   M76729
KEYWORDS    alpha-1 type V collagen.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 7138)
  AUTHORS   Greenspan,D.S., Cheng,W. and Hoffman,G.G.
  TITLE     The pro-alpha 1(V) collagen chain. Complete primary structure,
            distribution of expression, and comparison with the pro-alpha 1(XI)
            collagen chain
  JOURNAL   J. Biol. Chem. 266, 24727-24733 (1991)
  STANDARD  full automatic
COMMENT     NCBI gi: 189519
FEATURES             Location/Qualifiers
     source          1..7138
                     /organism="Homo sapiens"
                     /sequenced_mol="cDNA to mRNA"
     CDS             230..5746
                     /gene="COL5A1"
                     /map="9q34.2-q34.3"
                     /note="NCBI gi: 189520"
                     /codon_start=1
                     /gdb_xref="G00-131-457"
                     /product="pro-alpha-1 type V collagen"
                     /translation="MDVHTRWKARSALRPGAPLLPPLLLLLLWAPPPSRAAQPADLLK
                     VLDFHNLPDGITKTTGFCATRRSSKGPDVAYRVTKDAHVSAPTKQLYPASAFPEDFSI
                     LTTVKAKKGSQAFLVSIYNEQGIQQIGLELGRSPVFLYEDHTGKPGPEDYPLFRGINL
                     SDGKWHRIALSVHKKNVTLILDCKKKTTKFLDRSDHPMIDINGIIVFGTRILDEEVFE
                     GDIQQLLFVSDHRAAYDYCEHYSPDCDTAVPDTPQSQDPNPDEYYTEGDGEGETYYYE
                     YPYYEDPEDLGKEPTPSKKPVEAAKETTEVPEELTPTPTEAAPMPETSEGAGKEEDVG
                     IGDYDYVPSEDYYTPSPYDDLTYGEGEENPDQPTDPGAGAEIPTSTADTSNSSNPRPP
                     PGEGADDLEGEFTEETIRNLDENYYDPYYDPTSSPSEIGPGMPANQDTIYEGIGGPRG
                     EKGQKGEPAIIEPGMLIEGPPGPEGPAGLPGPPGTMGPTGQVGDPGERGPPGRPGLPG
```



```
                     ADGLPGPPGTMLMLPFRFGGGGDAGSKGPMVSAQESQAQAILQQARLALRGPAGPMGL
                     TGRPGPVGPPGSGGLKGEPGDVGPQGPRGVQGPPGPAGKPGRRGRAGSDGARGMPGQT
                     GPKGDRGFDGLAGLPGEKGHRGDPGPSGPPGPPGDDGERGDDGEVGPRGLPGEPGPRG
                     LLGPKGPPGPPGPPGVTGMDGQPGPKGNVGPQGEPGPPGQQGNPGAQGLPGPQGAIGP
                     PGEKGPLGKPGLPGMPGADGPPGHPGKEGPPGEKGGQGPPGPQGPIGYPGPRGVKGAD
                     GIRGLKGTKGEKGEDGFPGFKGDMGIKGDRGEIGPPGPRGEDGPEGPKGRGGPNGDPG
                     PLGPPGEKGKLGVPGLPGYPGRQGPKGSIGFPGFPGANGEKGGRGTPGKPGPRGQRGP
                     TGPRGERGPRGITGKPGPKGNSGGDGPAGPPGERGPNGPQGPTGFPGPKGPPGPPGKD
                     GLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGERGHPGPPGPPGEQGLPG
                     LAGKEGTKGDPGPAGLPGKDGPPGLRGFPGDRGLPGPVGALGLKGNEGPPGPPGPAGS
                     PGERGPAGAAGPIGIPGRPGPQGPPGPAGEKGAPGEKGPQGPAGRDGLQGPVGLPGPA
                     GPVGPPGEDGDKGEIGEPGQKGSKGDKGEQGPPGPTGPQGPIGQPGPSGADGEPGPRG
                     QQGLFGQKGDEGPRGFPGPPGPVGLQGLPGPPGEKGETGDVGQMGPPGPPGPRGPSGA
                     PGADGPQGPPGGIGNPGAVGEKGEPGEAGEPGLPGEGGPPGPKGERGEKGESGPSGAA
                     GPPGPKGPPGDDGPKGSPGPVGFPGDPGPPGEPGPAGQDGPPGDKGDDGEPGQTGSPG
                     PTGEPGPSGPPGKRGPPGPAGPEGRQGEKGAKGEAGLEGPPGKTGPIGPQGAPGKPGP
                     DGLRGIPGPVGEQGLPGSPGPDGPPGPMGPPGLPGLKGDSGPKGEKGHPGLIGLIGPP
                     GEQGEKGDRGLPGPQGSSGPKGEQGITGPSGPIGPPGPPGLPGPPGPKGAKGSSGPTG
                     PKGEAGHPGPPGPPGPPGEVIQPLPIQASRTRRNIDASQLLDDGNGENYVDYADGMEE
                     IFGSLNSLKLEIEQMKRPLGTQQNPARTCKDLQLCHPDFPDGEYWVDPNQGCSRDSFK
                     VYCNFTAGGSTCVFPDKKSEGARITSWPKENPGSWFSEFKRGKLLSYVDAEGNPVGVV
                     QMTFLRLLSASAHQNVTYHCYQSVAWQDAATGSYDKALRFLGSNDEEMSYDNNPYIRA
                     LVDGCATKKGYQKTVLEIDTPKVEQVPIADIMFNDFGEASQKFGFEVGPACFMG"
BASE COUNT     1546 a   2238 c   2278 g   1076 t
ORIGIN

Initial Score    =    139  Optimized Score  =    399  Significance =  8.57
Residue Identity =    50%  Matches          =    473  Mismatches   =   347
Gaps             =    108  Conservative Substitutions               =     0

           X          10         20         30         40         50
           GTCGACTG---GAACGAGACGA-CCTGCTCCAGACCATTCAGCAATC-CTGCTC-TGGGCTC
           | | | ||   || || | | | |||| ||   || |  | |   |  || |  |||||
   GGAGAAACGGGGCCCAATGGGTGAGCGTGGCCACCCTGGGCC---CCCTGGACCCCCCGGTGAACAGGGGCTT
     3250    X 3260       3270       3280       3290       3300       3310

     60         70         80         90        100        110        120
   CCAAGGGC--TGAGATTACAGGTGTAACCGAGATAAGACCCTGG--CTG-AGTCAACGATGCACAAGGCTCG
   ||  |||| ||     | ||  |  | |||     ||||| ||  ||| || |  |  ||   ||| |
   CC--GGGCCTTGCTGGAAAAGAAGGGA-CGAAGGGTGACCCAGGCCCTGCAGGCCTCCCTGGGAAAGACGGC
        3320       3330       3340       3350       3360       3370       3380

      130        140        150        160        170        180        190
   AATCCTAGGGATGGGAGG---GGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCC--TTCTTGCCCTTTGCCA
     ||| |||  |  | ||       ||| | || |||     ||| ||   |||  |    || ||| | |
   CCTCC-AGGATTACGTGGTTTCCCTGGGGA-CCGAGG-----GCTTCCTGGTCCAGTGGGAGCTCTTGGACT
        3390       3400       3410             3420       3430       3440

          200        210        220        230        240        250        260
   GCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGGAGCAACAGCTGCCT
   |  |||||       |||  || |  ||||||     | || | |  |||  |   |||   ||||||
   G-AAAGGCAATGAAGGGCCCCCTGGCCCACCAGGCCCTGCGGGATCTCCAGGGGAGAGAGGTCCAGCTGGAG
   3450       3460       3470       3480       3490       3500       3510

        270        280        290        300        310        320        330
   GGGCCGAGGCTCTGGGAGCTCC-GTGGCTTCTAGGGCTATGGGACTCCTTGGG--GGCAGCAGGGTGGAGGA
     || | |  | |||  ||| | |    || || |   ||||| ||  |||  ||||| || |    ||
   CCGCTGGGCCCATCGGAATTCCAGGGAGACCTGGGCCCCAGGGACCCCCAGGGCCGGCAGGAGAGAAAGGGG
 3520       3530       3540       3550       3560       3570       3580       3590
```

```
        340       350       360             370       380       390
GTCC-GGCTGGCACTGCACCTCCAGGCACCG-----AGAGAAATTCTGGCAGGCCTTCCCGATACAGGGCTT
 ||| |||  | |  || || | |||| | |      ||||   |||  ||||    ||   |  |||||
CTCCTGGCGAGAAAGGC-CCACAAGGCCCAGCTGGCCGAGACGGTCT-CCAGG--GGCC---TGTGGGGCTC
      3600      3610      3620      3630      3640       3650

400       410       420       430       440       450       460
CAGAG-CAAGC-AGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGGTCTGGACGAATCGCAGACAT
| | | |||  ||  ||||  |||  || |   || |  ||||||   ||    || | ||||  ||
CCGGGTCCAGCTGGCCCTGTG--GGTCCCCCTGGAGAAGACGGAGATAAGGG---AGA-G-ATCGGGGA---
 3660      3670      3680      3690      3700       3710

470       480       490       500       510       520       530
TCTGGTAGGGGCTGGAAGGTA-CATGAGGTGACAAAATGTATCTCGGT-GTTGACGTCCTCCAGAAGC----
    |  |||||| | |||| | || | || ||||||    | |  |||  |    | ||| |||   |
----GCCGGGGGCAGAAAGGAAGCAAG-GGGGACAAAGGAGAACAGGGTCCTCCTGGGCCTACAGGTCCTCAA
   3720      3730      3740      3750      3760      3770      3780

    540       550       560       570       580       590       600
GTTTGCATCTTAGACCCTGCCACAGTCTTCAGTTG--CTCTATCCAG---CGCTGGGCTAGGAAGAGGCTCC
|    ||||  | | || | | |  |||  || ||    | | || |   | | |||| | | || | ||
GGCCCCATCGGACAGCCAGGC-CCCTCTGGAGCTGACGGCGAGCCGGGGCCTCGGGGCCA-GCAGGGCCTTT
  3790      3800      3810      3820      3830      3840      3850

       610       620       630       640       650       660
ACAAGGCCTTGCAGTGCTTCTCGT-CCTGAAGATTGAC-GGCCACAGTGACTGGGTAATCTTTAAGCAGG--
 |  |||       ||| |    || || || | ||  | || | |  ||   ||  | |      |||||
TC-GGGCAGAAAGGTGATGAAGGTCCCAGAGGCTTTCCTGGACCCCCTG---GGCCAGTGGGGCTGCAGGGT
   3860      3870      3880      3890      3900       3910      3920

670       680       690       700       710       720       730
TGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGAC-TGTGGCTGAAGTAACAGTCAGG
| | |||   || ||| |  |        ||||   || || | | ||| || | | || |   |  | ||
TTGCCAG--GAC-CTCCAGGC------GAGAAGGGTGA-GACAGGAGACGTGGGCCAGATGGGCCCCCCGGG
      3930      3940      3950      3960      3970      3980

740       750       760       770       780       790       800
TGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGGAATTTGGGCTCCAGGCTGG---
|  |||  ||   | |||||  |  | | |||  |||  ||   | ||   |    |  | ||||| |||
TCCCCCTGGCCCCCGAGGAC-CCTCCGGAGCTCCAGGTGC-TGATGGCCCACAAGGTCCCCCAGG-TGGAAT
  3990      4000      4010      4020      4030      4040      4050

810       820       830       840       850       860       870
CGCCAGCACTGTCATCTCTGCCGGGGACCCCTCATGCCTGTGACAGAGCAGGT-CGTCT--CGTTCCAGTCG
 |  | | |||    |  ||   | ||      |  ||||| |  | |||||||  | ||  | |||| |  |
AGGAAACCCTG-GTGCAGTGGGAGAGA-AGGGCGAGCCTG-GCGA-AGCAGGTGAGCCTGGCCTTCCGGGAG
   4060      4070      4080      4090      4100      4110      4120

 X
AC
|
AAGGCGGCCCCC
 X   4130
```

11. US-08-162-407-1' (1-879)
    HUMPA1V Human pro-alpha-1 (V) collagen mRNA, complete cds.

LOCUS HUMPA1V 7138 bp ss-mRNA PRI 14-JAN-1992
DEFINITION Human pro-alpha-1 (V) collagen mRNA, complete cds.

ACCESSION M76729
KEYWORDS alpha-1 type V collagen.
SOURCE Homo sapiens cDNA to mRNA.
  ORGANISM Homo sapiens
  Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
  Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE 1 (bases 1 to 7138)
  AUTHORS Greenspan,D.S., Cheng,W. and Hoffman,G.G.
  TITLE The pro-alpha-1(V) collagen chain: Complete primary structure,
  distribution of expression, and comparison with the pro-alpha-1(XI)
  collagen chain
  JOURNAL J. Biol. Chem. 266, 24727-24733 (1991)
  STANDARD full automatic
COMMENT NCBI gi: 189519
FEATURES Location/Qualifiers
  source 1..7138
  /organism="Homo sapiens"
  /sequenced_mol="cDNA to mRNA"
  CDS 230..5746
  /gene="COL5A1"
  /map="Unassigned"
  /note="NCBI gi: 189520"
  /codon_start=1
  /product="pro-alpha-1 type V collagen"
  /translation="MDVHTRWKARSALRPGAPLLPPLLLLLLWAPPPSRAAQPADLLK
  VLDFHNLPDGITKTTGFCATRRSSKGPDVAYRVTKDAHVSAPTKQLYPASAFPEDFSI
  LTTVKAKKGSQAFLVSIYNEQGIQQIGLELGRSPVFLYEDHTGKPGPEDYPLFRGINL
  SDGKWHRIALSVHKKNVTLILDCKKKTTKFLDRSDHPMIDINGIIVFGTRILDEEVFE
  GDIQQLLFVSDHRAAYDYCEHYSPDCDTAVPDTPQSQDPNPDEYYTEGDGEGETYYYE
  YPYYEDPEDLGKEPTPSKKPVEAAKETTEVPEELTPTPTEAAPMPETSEGAGKEEDVG
  IGDYDYVPSEDYYTPSPYDDLTYGEGEENPDQPTDPGAGAEIPTSTADTSNSSNPRPP
  PGEGADDLEGEFTEETIRNLDENYYDPYYDPTSSPSEIGPGMPANQDTIYEGIGGPRG
  EKGQKGEPAIIEPGMLIEGPPGPEGPAGLPGPPGTMGPTGQVGDPGERGPPGPRGPLPG
  ADGLPGPPGTMLMLPFRFGGGGDAGSKGPMVSAQESQAQAILQQARLALRGPAGPMGL
  TGRPGPVGPPGSGGLKGEPGDVGPQGPRGVQGPPGPAGKPGRRGRAGSDGARGMPGQT
  GPKGDRGFDGLAGLPGEKGHRGDPGPSGPPGPPGDDGERGDDGEVGPRGLPGEPGPRG
  LLGPKGPPGPPGPPGVTGMDGQPGPKGNVGPQGEPGPPGQQGNPGAQGLPGPQGAIGP
  PGEKGPLGKPGLPGMPGADGPPGHPGKEGPPGEKGGQGPPGPQGPIGYPGPRGVKGAD
  GIRGLKGTKGEKGEDGFPGFKGDMGIKGDRGEIGPPGPRGEDGPEGPKGRGGPNGDPG
  PLGPPGEKGKLGVPGLPGYPGRQGPKGSIGFPGFPGANGEKGGRGTPGKPGPRGQRGP
  TGPRGERGPRGITGKPGPKGNSGGDGPAGPPGERGPNGPQGPTGFPGPKGPPGPPGKD
  GLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGERGHPGPPGPPGEQGLPG
  LAGKEGTKGDPGPAGLPGKDGPPGLRGFPGDRGLPGPVGALGLKGNEGPPGPPGPAGS
  PGERGPAGAAGPIGIPGRPGPQGPPGPAGEKGAPGEKGPQGPAGRDGLQGPVGLPGPA
  GPVGPPGEDGDKGEIGEPGQKGSKGDKGEQGPPGPTGPQGPIGQPGPSGADGEPGPRG
  QQGLFGQKGDEGPRGFPGPPGPVGLQGLPGPPGEKGETGDVGQMGPPGPPGPRGPSGA
  PGADGPQGPPGGIGNPGAVGEKGEPGEAGEPGLPGEGGPPGPKGERGEKGESGPSGAA
  GPPGPKGPPGDDGPKGSPGPVGFPGDPGPPGEPGPAGQDGPPGDKGDDGEPGQTGSPG
  PTGEPGPSGPPGKRGPPGPAGPEGRQGEKGAKGEAGLEGPPGKTGPIGPQGAPGKPGP
  DGLRGIPGPVGEQGLPGSPGPDGPPGPMGPPGLPGLKGDSGPKGEKGHPGLIGLIGPP
  GEQGEKGDRGLPGPQGSSGPKGEQGITGPSGPIGPPGPPGLPGPPGPKGAKGSSGPTG
  PKGEAGHPGPPGPPGPPGEVIQPLPIQASRTRRNIDASQLLDDGNGENYVDYADGMEE
  IFGSLNSLKLEIEQMKRPLGTQQNPARTCKDLQLCHPDFPDGEYWVDPNQGCSRDSFK
  VYCNFTAGGSTCVFPDKKSEGARITSWPKENPGSWFSEFKRGKLLSYVDAEGNPVGVV
  QMTFLRLLSASAHQNVTYHCYQSVAWQDAATGSYDKALRFLGSNDEEMSYDNNPYIRA
  LVDGCATKKGYQKTVLEIDTPKVEQVPIADIMFNDFGEASQKFGFEVGPACFMG"
BASE COUNT 1546 a 2238 c 2278 g 1076 t
ORIGIN

Initial Score = 139 Optimized Score = 399 Significance = 8.57

```
Residue Identity =    50%  Matches          =    473  Mismatches   =   347
Gaps             =    108  Conservative Substitutions              =     0

            X          10         20         30         40         50
            GTCGACTG---GAACGAGACGA-CCTGCTCCAGACCATTCAGCAATC-CTGCTC-TGGGCTC
            | | | ||   || || | | | ||||  ||   || |  | |    |  ||  |  |||||
    GGAGAAACGGGCCCAATGGGTGAGCGTGGCCACCCTGGGCC---CCCTGGACCCCCCGGTGAACAGGGGCTT
      3250    X 3260      3270       3280       3290       3300       3310

      60         70         80         90         100        110        120
    CCAAGGGC--TGAGATTACAGGTGTAACCGAGATAAGACCCTGG--CTG-AGTCAACGATGCACAAGGCTCG
    ||  ||||  ||     | ||  |  | |||     ||||| ||  ||| || |  |  ||   ||| |
    CC--GGGCCTTGCTGGAAAAGAAGGGA-CGAAGGGTGACCCAGGCCCTGCAGGCCTCCCTGGGAAAGACGGC
         3320       3330       3340       3350       3360       3370       3380

        130        140        150        160        170        180        190
    AATCCTAGGGATGGGAGG---GGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCC--TTCTTGCCCTTTGCCA
      ||| |||  |  | ||        ||| | || |||     ||| ||   |||  |    || ||| | |
    CCTCC-AGGATTACGTGGTTTCCCTGGGGA-CCGAGG-----GCTTCCTGGTCCAGTGGGAGCTCTTGGACT
         3390       3400       3410           3420       3430       3440

          200        210        220        230        240        250        260
    GCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCAACAGCTGCCT
    |  |||||       |||  || |   ||||||     | || | |  |||  |   |||   ||||||
    G-AAAGGCAATGAAGGGCCCCCTGGCCCACCAGGCCCTGCGGGATCTCCAGGGGAGAGAGGTCCAGCTGGAG
    3450       3460       3470       3480       3490       3500       3510

         270        280        290        300        310        320        330
    GGGCCGAGGCTCTGGGAGCTCC-GTGGCTTCTAGGGCTATGGGACTCCTTGGG--GGCAGCAGGGTGGAGGA
      || | | |  | |||  ||| | |    || || |  |||||  ||  |||  ||||| || |    ||
    CCGCTGGGCCCATCGGAATTCCAGGGAGACCTGGGCCCCAGGGACCCCCAGGGCCGGCAGGAGAGAAAGGGG
  3520       3530       3540       3550       3560       3570       3580       3590

           340        350        360            370        380        390
    GTCC-GGCTGGCACTGCACCTCCAGGCACCG------AGAGAAATTCTGGCAGGCCTTCCCGATACAGGGCTT
     ||| |||  | |  || || | |||| | |      ||||  |||  ||||    ||   |  |||||
    CTCCTGGCGAGAAAGGC-CCACAAGGCCCAGCTGGCCGAGACGGTCT-CCAGG--GGCC---TGTGGGGCTC
          3600       3610       3620       3630       3640           3650

    400        410        420        430        440        450        460
    CAGAG-CAAGC-AGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGGTCTGGACGAATCGCAGACAT
    | | | | |||  ||  ||||  |||  || |   || |  ||||||   ||    || | ||||  ||
    CCGGGTCCAGCTGGCCCTGTG--GGTCCCCCTGGAGAAGACGGAGATAAGGG---AGA-G-ATCGGGGA---
     3660       3670       3680       3690       3700           3710

    470        480        490        500        510        520        530
    TCTGGTAGGGGCTGGAAGGTA-CATGAGGTGACAAAATGTATCTCGGT-GTTGACGTCCTCCAGAAGC----
        |  |||||  | ||||  | ||  | ||  ||||||    | |  |||  |    | ||| |||   |
    ----GCCGGGGCAGAAAGGAAGCAAG-GGGGACAAAGGAGAACAGGGTCCTCCTGGGCCTACAGGTCCTCAA
       3720       3730       3740       3750       3760       3770       3780

        540        550        560        570        580        590        600
    GTTTGCATCTTAGACCCTGCCACAGTCTTCAGTTG--CTCTATCCAG---CGCTGGGCTAGGAAGAGGCTCC
    |    ||||  | | || | | |  |||  || ||    | | || |   | | |||| | | || | ||
    GGCCCCATCGGACAGCCAGGC-CCCTCTGGAGCTGACGGCGAGCCGGGGGCCTCGGGGCCA-GCAGGGCCTTT
      3790       3800       3810       3820       3830       3840       3850

           610        620        630        640        650        660
    ACAAGGCCTTGCAGTGCTTCTCGT-CCTGAAGATTGAC-GGCCACAGTGACTGGGTAATCTTTAAGCAGG--
      |  |||       ||| |    || || || | ||  | || | |  ||   ||  | |      |||||
```



```
    TC-GGGCAGAAAGGTGATGAAGGTCCCAGAGGCTTTCCTGGACCCCCTG---GGCCAGTGGGGCTGCAGGGT
       3860       3870       3880       3890       3900           3910       3920

    670        680        690        700        710        720        730
    TGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGAC-TGTGGCTGAAGTAACAGTCAGG
    | | |||    || ||| |  |          ||||   || || | | ||| || | | | || |    |   | ||
    TTGCCAG--GAC-CTCCAGGC------GAGAAGGGTGA-GACAGGAGACGTGGGCCAGATGGGCCCCCCGGG
          3930       3940            3950       3960       3970       3980

   740        750        760        770        780        790        800
    TGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGGAATTTGGGCTCCAGGCTGG---
    |  |||  ||   | ||||| |  | | |||  |||  ||   | ||   |    |  | ||||| |||
    TCCCCCTGGCCCCCGAGGAC-CCTCCGGAGCTCCAGGTGC-TGATGGCCCACAAGGTCCCCCAGG-TGGAAT
      3990       4000       4010       4020       4030       4040       4050

    810        820        830        840        850        860        870
    CGCCAGCACTGTCATCTCTGCCGGGGACCCCTCATGCCTGTGACAGAGCAGGT-CGTCT--CGTTCCAGTCG
     |  | | |||    |  ||   | ||     |  ||||| |  | |||||||  | ||  | |||| |  |
    AGGAAACCCTG-GTGCAGTGGGAGAGA-AGGGCGAGCCTG-GCGA-AGCAGGTGAGCCTGGCCTTCCGGGAG
      4060       4070       4080       4090       4100       4110       4120

     X
    AC
    |
    AAGGCGGCCCCC
     X   4130
```

12. US-08-162-407-1' (1-879)
    HS4ULIR3 epstein-barr virus simple repeat array (ir3).

```
LOCUS       HS4ULIR3     1150 bp ds-DNA             VRL       15-JUN-1990
DEFINITION  epstein-barr virus simple repeat array (ir3).
ACCESSION   J02079
KEYWORDS    repeat region.
SOURCE      ebv (epstein barr virus) from human.
  ORGANISM  Epstein-Barr virus
            Viridae; ds-DNA enveloped viruses; Herpesviridae;
            Gammaherpesvirinae.
REFERENCE   1  (bases 1 to 1150)
  AUTHORS   Heller,M., van Santen,V.L. and Keiff,E.
  TITLE     simple repeat sequence in epstein-barr virus dna is transcribed in
            latent and productive infections
  JOURNAL   J. Virol. 44, 311-320 (1982)
  STANDARD  full automatic
COMMENT     NCBI gi: 330445
FEATURES             Location/Qualifiers
     source          1..1150
                     /organism="Epstein-Barr virus"
BASE COUNT      302 a    169 c    633 g     46 t
ORIGIN
```

```
Initial Score    =    134  Optimized Score  =    370  Significance =  8.18
Residue Identity =    49%  Matches          =    443  Mismatches   =   355
Gaps             =    102  Conservative Substitutions              =     0

              X       10         20         30         40         50         60
              GTCGACTGGAACGAG-ACGACCTGCTCCA-GACCATTCAGCAATCCTGCTCTGGGCTCCCAA
                 ||  || | ||| | ||    |   || || ||   || |           ||||    |
    GGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCA-GGAGGAGGGGCAGGAGGGGC-AGGAG
```

```
      300  X    310        320        330        340        350        360
        70         80         90        100        110        120
GGGCTGAGATTACAGGTGTA-ACCGAG--ATAAGACCCTG-GCTGAGTCA-ACGATGCACAAGGCTC-GAAT
GGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG
     370        380        390        400        410        420        430

 130        140        150        160        170        180        190
CCTAGGGATGGG-AGG-GGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCC--CTTTGCCAGCGA
CAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAG-GA
   440        450        460        470        480        490        500

       200        210        220        230        240        250
AGGCC----CCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCA---GCAGCAGCAGGAGCAACAGCTG
GGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
  510        520        530        540        550        560        570

260        270        280        290        300        310        320        330
CCTGGGC-CGAGGCTCTGGGAGCTCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAGGGTGGAGG
GAGGGGCAGGAGGGGC-AGGAGC-AGGAGG-----AGGGGCA--GGA-GCAGGAGGGGCAGGAGGG-GCAGG
  580        590        600            610        620        630

       340        350        360        370        380        390        400
AGTCCGGCTGGCACTGCACCTCCAGGCACCGAGAGAAATTCTGGCAGGCCTTCCCGATACAGG-GCTTCAGA
AG--GGGCAGGAGCAGGAGGGGCAGG-AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGG
   640        650        660        670        680        690        700

      410        420        430        440        450        460
GCAAGCAGCTG--TGTGCAGGTGTCCTTCAGGAGGTG--GGAG-ATGTTGGTCTGGACG---AATCGCA-GA
GGCAGGAGCAGGAGGGGCAGGAG-----CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGA
  710        720            730        740        750        760        770

   470        480            490        500        510        520
CATTCTGGTAGGGGCTGGA-----AGGTACATGAGGTG-ACAAAATGTATCTCGGTGTTGACGTCC---TCC
GGGGCAGG-AGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGC
     780        790        800        810        820        830        840

530        540        550        560        570        580        590
AGAAGCGTTTGCATCTTAGAC-CCTGC--CACAGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAGGAAGAGG
AGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA-GCAG-GAGGGGC-AGG-AGGGG
    850        860        870        880        890            900        910

 600        610        620        630        640        650        660
CTCCACAAGGCCTTGCA-GTGCTTCTCG-TCCTGAAGATTGACGGCCACAGTGACTGGGTAATCTTTAAGCA
CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG-GA-GGGGCA-----GGAGCA
      920        930        940        950        960            970

 670        680        690        700        710        720        730
GGTGGT-CAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAG--GAGATGGGACT-GTGGCTGAAGT--AAC
GGAGGTGGAGGCCGGGGTCGAGGAGGCA-GTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGG
```



```
     980        990       1000       1010       1020       1030       1040
        740        750        760        770        780        790        800
AGTCAGGTGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGGAATTTGGGCTCCAGG
 AGGTA-GTGGAGGCCGCCGG---GG-TAGAGGACGTGAAAGAGC--CA--GGGGGGGAAGTCGTGAAAGAGC
1050       1060           1070       1080           1090       1100       1110

     810        820        830        840          X       860
CTGGCGCCAGCACTGTCATCTCTGCCGGGGACCCCTCATGCCTGTGACAGAGCAGG
CAGG-GGGAG--AGGTCGT---GGACGTGGAGAAAAGAGGCCCATG
          1120       1130       1140       1150
```

13. US-08-162-407-1' (1-879)
    CHPINVOL Chimpanzee (P.paniscus) involucrin, complete cds.

```
LOCUS       CHPINVOL     1683 bp ds-DNA             PRI       15-MAR-1990
DEFINITION  Chimpanzee (P.paniscus) involucrin, complete cds.
ACCESSION   M26514
KEYWORDS    epidermal protein; involucrin.
SOURCE      P.paniscus (strain CHP) keratinocyte DNA.
  ORGANISM  Pan paniscus
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Pongidae.
REFERENCE   1  (bases 1 to 1683)
  AUTHORS   Djian,P. and Green,H.
  TITLE     Vectorial expansion of the involucrin gene and the relatedness of
            the hominoids
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86, 8447-8451 (1989)
  STANDARD  full automatic
COMMENT     Draft entry and computer readable copy of sequence [1] kindly
            submitted by P.Djian, 26-JUL-1989.

            NCBI gi: 176809
FEATURES             Location/Qualifiers
     source          1..1683
                     /organism="Pan paniscus"
     CDS             1..1683
                     /note="involucrin;  NCBI gi: 176810"
                     /codon_start=1
                     /translation="MSQQHTLPVTLSPALSQELLKTVPPPVNTQQEQMKQPTPLPPPC
                     QKMPVELPVEVPSKQEEKHMTAVKGLPEQECEQQQQEPQEQELQQQHWEQHEEYQKAE
                     NPEQQLKQEKAQRDPQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELPEQQE
                     GHLKHLEQREGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLELPEQQEGQLKHLEQ
                     QEGQLKHLEHQEGQLEVPEEQVGQLKYLEQQEGQLKHLDQQEKQPELPEQQVGQLKHL
                     EQQEGQPKHLEQQKGQLEHLEEQEGQLKHLEQQEGQLEHLEHQEGQLGLPEQQVQQLK
                     QLEKEEGQPKHLEEEEGQLKHLVQQEGQLEHLVQQEGQLEHLVQQEGQLEQQEGQVEH
                     LEQQVEHLEQLGQLKHLEEQEGQLKHLEQQQGQLGVPEQVGQPKNLEQEEKQLELPEQ
                     QEGQLKHLEKQEAQLELPEQQVGQPKHLEQQEKQLEPPEQQDGQLKHLEQQEGQLKDL
                     EQQKGQLEQPVFAPAPGQVQDIQSALPTKGEVLLPLEHQQQKQEVQWPPKHK"
BASE COUNT      496 a    419 c    597 g    171 t
ORIGIN

Initial Score    =    134  Optimized Score  =    407  Significance =  8.18
Residue Identity =    51%  Matches          =    493  Mismatches   =   341
Gaps             =    120  Conservative Substitutions               =     0

            X       10        20        30        40        50
```

```
            GTCGACTGGAACGAGACGACCTGCTCCA-GACCA--TTCAGCAATCCTGCTCTGGGCTCCCA
            | |  ||||| |     || | ||   | |  ||  |  |||   || |  | |
  CAGCAGAAGGGGCAGCTGGAGCTCCCAGAGCAGCAGGAGGGGCAGCTGGAGC-TCCCAGAGCAGCAGGAGGG
        570 X      580       590       600       610       620       630

 60        70        80                90       100       110       120
  AGGGCTGAGATTACAGGTGTAACCG--------AGAT-AAGACCCTGGCTGAGTCAACGATGCACAAGGCTCG
  |  |||||     | || | | | |        || | |||   ||||   |  | | || |  ||  ||| |
  ACAGCTGAAGCACCTGGAGCAGCAGGAGGGTCAGCTGAAGCATCTGGAGCA--CCAGGAGGGGCA--GCTGG
       640       650       660       670       680       690       700

      130       140       150       160       170       180
  A-ATCCTAG-GGATGGGAGGGGAGGGGCACCCCAGG---GTGGAGCTCC--CCCCTCCTTCT--TGC-CCTT
  |  ||| || |||  | |  || ||||||  |  | |    |||| |  |        |  ||   || |||
  AGGTCCCAGAGGA--GCA--GGTGGGGCAGCTGAAGTACCTGGAACAGCAGGAGGGGCAGCTGAAGCACCTG
         710       720       730       740       750       760

    190       200       210       220       230       240       250
  TGCCAGCGAAGGCCCCAGGCG--GCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGGAGCAACA
     ||||    |  ||| ||  ||| |||| |||| ||| ||  | ||||||  | |||| | ||||||  ||
  GATCAGCAGGAGAAGCAGCCGGAGCTCCCAG-AGCAGCAG-GT--GGGGCAGCTGAAGCACCTGGAGCAGCA
 770       780       790       800       810       820       830

    260       270       280       290       300       310       320
  GCTGCCTGGGCCGAGGC-TCTGGGAGCTCC----GTGGCTTCTAGGGC-TATGG-GACTCCTTGGGGGCAGC
  |  |    ||| | || ||| |||||  |    | |||  || | ||   ||| |   |    ||||||||
  GGAGGGGCAGCCTAAGCATCT-GGAGCAGCAGAAGGGGCAACTGGAGCACCTGGAGGAGCAGGAGGGGCAGC
   840       850       860       870       880       890       900

             330       340       350       360       370       380
  AG-----GGTGGAGGAGTCCGGCTGGCA-CTGCACCTCCAG--GCACCGAGAGAAATTCTGGCAGGCCTTCC
    |     ||||| ||   |  |||| ||| | | || |  ||||| || |||   | |   || || ||
  TGAAGCACCTGGAGCAGCAGGAGGGGCAGCTGGAGCACCTGGAGCACC-AG-GAAGGGCAGCTGGGGCTCCC
  910       920       930       940       950       960       970

      390       400       410       420       430       440
  CGA--TACAGG-GCTTCAGAGCAAGCAGCT-GTG-TGCAGG-TGTCCTTC---AGGAGGTGGGAGATGTTGG
   ||    |||| ||  |||   |||||||| | |  | ||| |  |  |  |    || |  | |||| |  ||
  AGAGCAGCAGGTGCAGCAGCTGAAGCAGCTAGAGAAGGAGGAGGGGCAGCCAAAGCACCT-GGAGGAGGAGG
  980       990      1000      1010      1020      1030      1040

  450       460       470       480       490       500       510
  TCTGGACGAATCGCAGACATTCTGGTAGGGGCTGGAAGGTACATGAGGTG-ACAAAATGTATCTCGGTGTTG
     ||   | | | ||  ||  ||||| |  || || |||  ||   || | ||    || | |  |  | |
  AGGGG--CAGTTGAAG-CA-CCTGGT-GCAGCAGG-AGGGGCAGCTGGAGCACCTGGTGCAGCAGGAGGGTC
1050        1060      1070      1080      1090      1100      1110

 520       530       540       550       560       570       580
  ACGTCCTCCAGAAGCGTTTGCATC-TTAGACCCTGCCACAGTCTTCAGTTGCTCTA--TCCAGC-GCTGGGC
  |    ||  || | |   |||| |   ||   | ||   || |  |||  |     |  | |||  ||||
  A---GCTGGAGCATCTGGTGCAGCAGGAGGGGCAGCTGGAG-CAGCAGGAGGGGCAGGTGGAGCACCTGGAG
       1120      1130      1140      1150      1160      1170      1180

   590       600       610           620       630       640       650
  TAGGAAGAG--GCTCCACAAGGCCTT---GCAG-TGCTTCTCGTCCTGAAGATTGACGGCCACAGTGACTGG
   || | |  || ||   ||   ||   |||| ||   | | |   | ||   || || ||     ||||
  CAGCAGGTGGAGCACCTGGAGCAGTTGGGGCAGCTGAAGCACCTAGAGGAGCAGGAGGGGCA-----ACTGA
      1190      1200      1210      1220      1230      1240

       660       670       680       690       700       710       720
```



```
    GTAATCTTTAAGCAGGTGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAG--ATGGGA-CTGTG
       ||| |  |||||  |  |||   ||  | |      || |   | || ||| | ||  |  | | |||
    AGCATC-TGGAGCAGCAG--CAG-GGACAGT-TGGGGGTCCCAGAGCAGGTGGGGCAGCCAAAGAACCTGGA
1250       1260          1270       1280       1290       1300       1310

          730       740       750           760       770       780
    GCTGAAGTAACAGTCAGGTGTCCCCCGCAG----GCAAG-GACTCAGCAGCAGCAACAGCAGCAACAGGGAG
    || | || |  || ||| ||   | | |||     ||||| |   |||| | |||| | | || | ||||  |
    GCAGGAGGAGAAG-CAGCTGGAGCTCCCAGAGCAGCAAGAGGGCCAGCTGAAGCACCTGGAGAAGCAGGAGG
    1320       1330      1340      1350      1360      1370      1380

   790           800       810       820       830       840       850
    GA-ATTTGG-GCT-CCAG-GCTGGCGCCAGCACTGTCATCTCTGCCGGGGACCCCTCATGCCTGTGACAGAG
     | |  ||| ||| |||| || |  |  || || | ||    |  | || |       || ||  |  |  ||
    CACAGCTGGAGCTCCCAGAGCAGCAGGTAGGACAGCCAAAGCACCTGGAG------CA-GCAGGAAAAGCAG
  1390       1400      1410      1420      1430              1440      1450

     860       870           X
    C-AGGTCGTCTCGTTC--CA-GTCG-AC
    | ||  |  |  |  |  || | || ||
    CTAGAGCCCCCAGAGCAGCAGGACGGACAACTAAAACA
        1460      1470      1480      1490
```

14. US-08-162-407-1' (1-879)
    U02454 Cloning vector pCMVEBNA, complete sequence.

```
LOCUS       U02454       5452 bp    DNA   circular  SYN       08-NOV-1993
DEFINITION  Cloning vector pCMVEBNA, complete sequence.
ACCESSION   U02454
KEYWORDS    .
SOURCE      Cloning vector pCMVEBNA.
  ORGANISM  Cloning vector pCMVEBNA
            Artificial sequences; Cloning vector.
REFERENCE   1  (bases 1 to 5452)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.1
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 5452)
  AUTHORS   Swirski,R.A., Van Den Berg,D., Murphy,A.J., Lambert,C.M.,
            Friedberg,E.C. and Schimke,R.T.
  TITLE     Improvements in the Epstein-Barr-based shuttle vector system for
            direct cloning in human tissue culture cells
  JOURNAL   Methods: A Companion to Methods in Enzymology 4, 133-142 (1992)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 5452)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            4030 Fabian Way, Palo Alto, CA 94303, USA
  STANDARD  full automatic
COMMENT     This vector can be obtained from CLONTECH Laboratories, Inc., 4030
            Fabian Way, Palo Alto, CA 94303, USA. To place an order call (415)
            424-8222 or (800) 662-2566, extension 1. International customers,
            please contact your local distributor.  For technical information,
            call (415) 424- 8222 or (800) 662-2566, extension 3.
            This sequence has been compiled from information in the sequence
            databases, published literature and other sources; this vector has
            not been completely sequenced. If you suspect there is an error in
```

this sequence, please contact CLONTECH's Technical Service Department at (415) 424-8222 or (800) 662-2566, extension 3 or E-mail CLONTECH@BIOTECHNET.COM.

NCBI gi: 413820

```
FEATURES             Location/Qualifiers
     source          1..5452
                     /organism="Cloning vector pCMVEBNA"
BASE COUNT     1362 a   1246 c   1736 g   1108 t
ORIGIN
```

Initial Score = 134 Optimized Score = 385 Significance = 8.18
Residue Identity = 48% Matches = 461 Mismatches = 379
Gaps = 108 Conservative Substitutions = 0

```
            X       10        20        30        40        50        60
            GTCGACTGGAACGAG-ACGACCTGCTCCA-GACCATTCAGCAATCCTGCTCTGGGCTCCCAA
                  ||  || | ||| | ||   |   || || ||   || |          ||||    |
   GGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCA-GGAGGAGGGGCAGGAGGGGC-AGGAG
    1090      1100      1110      1120      1130      1140      1150

          70        80        90       100       110       120
   GGGCTGAGATTACAGGTGTA-ACCGAG--ATAAGACCCTG-GCTGAGTCA-ACGATGCACAAGGCTC-GAAT
   |||| |      ||||  | |    |||    | || |  | |  | | ||   |  |||  |||  | | |
   GGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG
    1160      1170      1180      1190      1200      1210      1220

    130       140       150       160       170       180       190
   CCTAGGGATGGG-AGG-GGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCC--CTTTGCCAGCGA
   |    ||| ||| ||| | ||| | |      || || ||      |      | |     | ||| ||
   CAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAG-GA
  1230      1240      1250      1260      1270      1280      1290

         200       210       220       230       240       250
   AGGCC----CCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCA---GCAGCAGCAGGAGCAACAGCTG
    || |     |||| ||  |   |||| | |||   ||| |||||  |   |||| |||||||| |   || |
   GGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGCAGGAGCAGGAGGAGGGGCAG
 1300      1310      1320      1330      1340      1350      1360

 260       270       280       290       300       310       320       330
   CCTGGGC-CGAGGCTCTGGGAGCTCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAGGGTGGAGG
      ||||  ||||  |  |||||   | ||     ||||  |  |||  |    |||||||  |||| | |||
   GAGGGGCAGGAGGGGC-AGGAGC-AGGAGG-----AGGGGCA--GGA-GCAGGAGGGGCAGGAGGG-GCAGG
 1370      1380      1390            1400      1410      1420

         340       350       360       370       380       390       400
   AGTCCGGCTGGCACTGCACCTCCAGGCACCGAGAGAAATTCTGGCAGGCCTTCCCGATACAGG-GCTTCAGA
   ||   ||| ||  | | |    |||| | |  ||| |       | |||       |   |||| ||   ||
   AG--GGGCAGGAGCAGGAGGGGCAGG-AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGCAGGAGCAGGAGG
1430       1440      1450      1460      1470      1480      1490

        410       420       430       440       450       460
   GCAAGCAGCTG--TGTGCAGGTGTCCTTCAGGAGGTG--GGAG-ATGTTGGTCTGGACG---AATCGCA-GA
   |  || ||| |   | |||||  |     |||||||  |  |||| | |  || | ||| |   |   ||| ||
   GGCAGGAGCAGGAGGGGCAGGAG-----CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGA
 1500      1510      1520           1530      1540      1550      1560

     470       480           490       500       510       520
   CATTCTGGTAGGGGCTGGA-----AGGTACATGAGGTG-ACAAAATGTATCTCGGTGTTGACGTCC---TCC
       | || |||||| |||     |||  || |||| | |  |   | |    ||    || |  |     |
```



```
   GGGGCAGG-AGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGC
    1570      1580      1590      1600      1610      1620      1630

   530       540       550       560       570       580       590
   AGAAGCGTTTGCATCTTAGAC-CCTGC--CACAGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAGGAAGAGG
   || ||     |||     | |   ||  |  | |  ||  ||   |  ||| |  |||| ||| || ||
   AGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA-GCAG-GAGGGGC-AGG-AGGGG
   1640      1650      1660      1670      1680      1690      1700

    600       610       620       630       640       650       660
   CTCCACAAGGCCTTGCA-GTGCTTCTCG-TCCTGAAGATTGACGGCCACAGTGACTGGGTAATCTTTAAGCA
   |   |  |||    ||| | |   |  |  |  || ||  | | |   ||| ||  ||| |       ||||
   CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG-GA-GGGGCA-----GGAGCA
     1710      1720      1730      1740      1750      1760

    670       680       690       700       710       720       730
   GGTGGT-CAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAG--GAGATGGGACT-GTGGCTGAAGT--AAC
   || |||  || |     || |     ||  | || |  || |  |||  || | | | ||| |  ||  |
   GGAGGTGGAGGCCGGGGTCGAGGAGGCA-GTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGG
1770      1780      1790      1800      1810      1820      1830      1840

         740       750       760       770       780       790       800
   AGTCA--GGTGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAA--CAGGGAGGAATTTGGGCTC
   ||  |  || | || |||  |   |||||  | |  ||| |  |  | ||  |  | | ||    |||
   AGGTAGTGGAGGCCGCCGGGGTAGAGGAC-GTGAAAGAGCCAGGGGGGGAAGTCGTGAAAGAGCCAGGGGGA
        1850      1860      1870      1880      1890      1900      1910

                 810       820       830       840       850
   CAGG--------CTGGCGCCA----GCACTGTCATCTCTGCCG-GGGAC-CCCTCATGCCTGTGACAGAGCA
    |||         |||  |   |   || | |   || | |  |   | |  | |||| |  ||  |  | |
   GAGGTCGTGGACGTGGAGAAAAGAGGCCCAGGAGTCCCAGTAGTCAGTCATCATCATCCGGGTCTC-CACCG
        1920      1930      1940      1950      1960      1970      1980

  860       870       X
   GGTCGTCTCGTTCCA-GTCGAC
    |  | | |  |||| || ||
   CGCAGGCCCCCTCCAGGTAGAAGGCCATTTTT
       1990      2000   X  2010
```

15. US-08-162-407-1' (1-879)
    U02428 Cloning vector pDR2, complete sequence.

```
LOCUS       U02428      10737 bp    DNA   circular  SYN       08-NOV-1993
DEFINITION  Cloning vector pDR2, complete sequence.
ACCESSION   U02428
KEYWORDS    .
SOURCE      Cloning vector pDR2.
  ORGANISM  Cloning vector pDR2
            Artificial sequences; Cloning vector.
REFERENCE   1  (bases 1 to 10737)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.1
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 10737)
  AUTHORS   Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
  TITLE     cDNA expression cloning in human cells using the plambdaDR2
            episomal vector system
  JOURNAL   Methods:A Companion to Methods in Enzymology 4, 111-131 (1992)
```

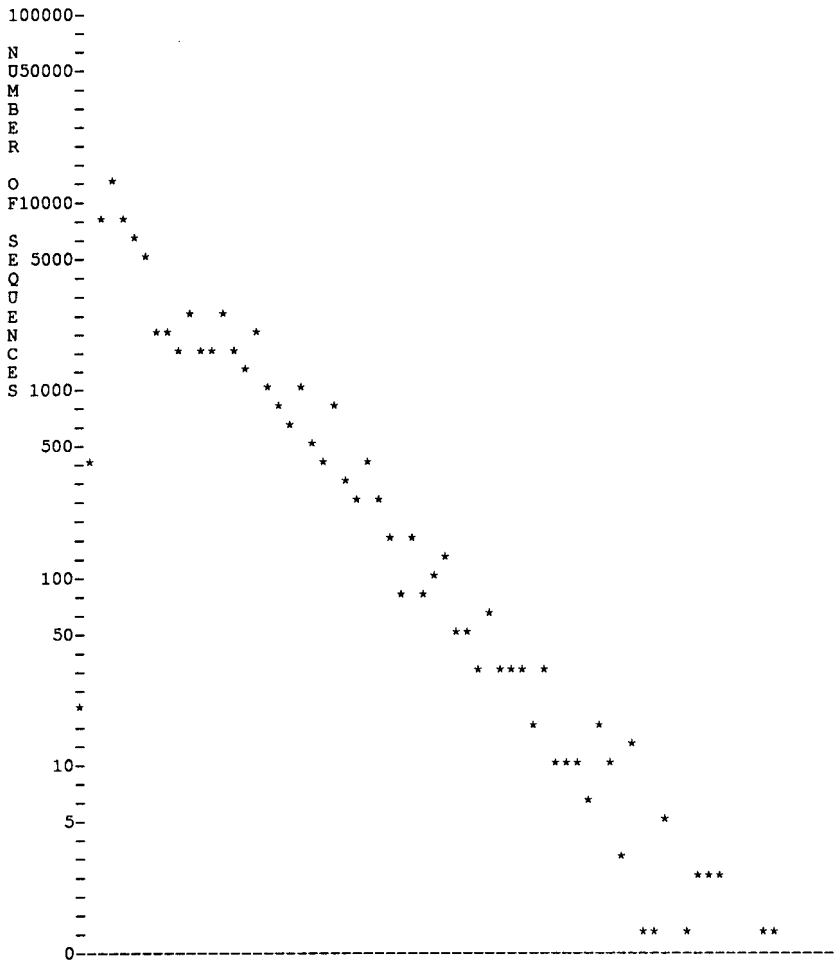
```
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 10737)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            4030 Fabian Way, Palo Alto, CA 94303, USA
  STANDARD  full automatic
COMMENT     This vector can be obtained from CLONTECH Laboratories, Inc., 4030
            Fabian Way, Palo Alto, CA 94303, USA. To place an order call (415)
            424-8222 or (800) 662-2566, extension 1. International customers,
            please contact your local distributor.  For technical information,
            call (415) 424- 8222 or (800) 662-2566, extension 3. This sequence
            was compiled by Andrew Murphy and revised at CLONTECH; this vector
            has not been completely sequenced. If you suspect there is an error
            in this sequence, please contact CLONTECH's Technical Service
            Department at (415) 424-8222 or (800) 662-2566, extension 3 or
            E-mail CLONTECH@BIOTECHNET.COM.

            NCBI gi: 413794
FEATURES            Location/Qualifiers
     source         1..10737
                    /organism="Cloning vector pDR2"
BASE COUNT     2686 a   2596 c   3033 g   2422 t
ORIGIN

Initial Score     =    134  Optimized Score  =    385  Significance =  8.18
Residue Identity  =    48%  Matches          =    461  Mismatches   =   379
Gaps              =    108  Conservative Substitutions               =     0

             X       10        20        30        40        50        60
             GTCGACTGGAACGAG-ACGACCTGCTCCA-GACCATTCAGCAATCCTGCTCTGGGCTCCCAA
                 ||  || | ||| | ||    |   || || ||   || |          ||||   |
   GGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCA-GGAGGAGGGGCAGGAGGGGC-AGGAG
 6620      6630      6640      6650      6660      6670      6680

          70        80        90       100       110       120
   GGGCTGAGATTACAGGTGTA-ACCGAG--ATAAGACCCTG-GCTGAGTCA-ACGATGCACAAGGCTC-GAAT
   |||| |      |||| | |    |||    | || |  | |  | | ||   | |||  |||  | | |
   GGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG
 6690      6700      6710      6720      6730      6740      6750      6760

    130       140       150       160       170       180       190
   CCTAGGGATGGG-AGG-GGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCC--CTTTGCCAGCGA
   |    ||| ||| ||| | ||| | |       || || ||      |         | |      | ||| ||
   CAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAG-GA
        6770      6780      6790      6800      6810      6820      6830

         200       210       220       230       240       250
   AGGCC----CCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCA---GCAGCAGCAGGAGCAACAGCTG
    || |    |||| ||  |   |||| | |||   ||| |||||  |   |||| |||||||| |   || |
   GGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
        6840      6850      6860      6870      6880      6890      6900

 260       270       280       290       300       310       320       330
   CCTGGGC-CGAGGCTCTGGGAGCTCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAGGGTGGAGG
      ||||  ||||  |  |||||   | ||     ||||  |  |||  |    |||||| |||| | |||
   GAGGGGCAGGAGGGGC-AGGAGC-AGGAGG-----AGGGGCA--GGA-GCAGGAGGGGCAGGAGGG-GCAGG
        6910      6920            6930      6940      6950      6960

          340       350       360       370       380       390       400
```



```
   AGTCCGGCTGGCACTGCACCTCCAGGCACCGAGAGAAATTCTGGCAGGCCTTCCCGATACAGG-GCTTCAGA
   ||   ||| ||  | | |    |||| | |  ||| |       | |||       |  |||| ||   ||
   AG--GGGCAGGAGCAGGAGGGGCAGG-AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGG
         6970      6980      6990      7000      7010      7020      7030

          410       420       430       440       450       460
   GCAAGCAGCTG--TGTGCAGGTGTCCTTCAGGAGGTG--GGAG-ATGTTGGTCTGGACG---AATCGCA-GA
   |  || ||| |   | |||||  |     |||||||  | |||| | |  || | ||| |   |   ||| ||
   GGCAGGAGCAGGAGGGGCAGGAG-----CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGA
        7040      7050      7060      7070      7080      7090

      470       480       490       500       510       520
   CATTCTGGTAGGGGCTGGA-----AGGTACATGAGGTG-ACAAAATGTATCTCGGTGTTGACGTCC---TCC
       | || |||||| |||     |||  || |||| | |   | |    ||   || | |     |
   GGGGCAGG-AGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGC
 7100      7110      7120      7130      7140      7150      7160

   530       540       550       560       570       580       590
   AGAAGCGTTTGCATCTTAGAC-CCTGC--CACAGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAGGAAGAGG
   || ||     |||     | |    ||   | | |   ||  ||    |  |||  |  |||| ||| || ||
   AGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA-GCAG-GAGGGGC-AGG-AGGGG
 7170      7180      7190      7200      7210      7220      7230

    600       610       620       630       640       650       660
   CTCCACAAGGCCTTGCA-GTGCTTCTCG-TCCTGAAGATTGACGGCCACAGTGACTGGGTAATCTTTAAGCA
   |   |  |||    ||| | |   | | | | || ||  | | |   ||| ||  ||| |       ||||
   CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG-GA-GGGGCA-----GGAGCA
 7240      7250      7260      7270      7280      7290            7300

    670       680       690       700       710       720       730
   GGTGGT-CAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAG--GAGATGGGACT-GTGGCTGAAGT--AAC
   || |||  || |     || |     ||  | || |  || |  |||  || | | | ||| |  ||  |
   GGAGGTGGAGGCCGGGGTCGAGGAGGCA-GTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGG
      7310      7320      7330      7340      7350      7360      7370

         740       750       760       770       780       790       800
   AGTCA--GGTGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAA--CAGGGAGGAATTTGGGCTC
   ||  |  || | || ||| |  |||||   | |  ||| |  | |  | | || |  |||
   AGGTAGTGGAGGCCGCCGGGGTAGAGGAC-GTGAAAGAGCCAGGGGGGGAAGTCGTGAAAGAGCCAGGGGGA
      7380      7390      7400      7410      7420      7430      7440

               810       820       830       840       850
   CAGG--------CTGGCGCCA----GCACTGTCATCTCTGCCG-GGGAC-CCCTCATGCCTGTGACAGAGCA
    |||         ||| |  |    || | |  || | |   | | | ||||  |  ||  |  | |
   GAGGTCGTGGACGTGGAGAAAAGAGGCCCAGGAGTCCCAGTAGTCAGTCATCATCATCCGGGTCTC-CACCG
     7450      7460      7470      7480      7490      7500      7510

   860       870       X
    GGTCGTCTCGTTCCA-GTCGAC
     |  | | |  |||| || ||
    CGCAGGCCCCCTCCAGGTAGAAGGCCATTTTT
     7520      7530      7540
> O <
O|  |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq1nsq.res made by  on Wed 5 Apr 95 17:37:49-PDT.
```

Query sequence being compared:US-08-162-407-1 (1-879)
Number of sequences searched: 57621
Number of scores above cutoff: 4354

Results of the initial comparison of US-08-162-407-1 (1-879) with:
Data bank : N-GeneSeq 17, all entries




```
SCORE 0    16   33   49   65   82   98   114  131  147
STDEV      0    1    2    3    4    5    6    7    8
```

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	18	13	14.85

Times:	CPU	Total Elapsed
	00:03:47.03	00:03:47.00

Number of residues: 24347505
Number of sequences searched: 57621
Number of scores above cutoff: 4354

Cut-off raised to 11.
Cut-off raised to 18.
Cut-off raised to 26.
Cut-off raised to 32.
Cut-off raised to 37.
Cut-off raised to 42.
Cut-off raised to 46.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 8 standard deviations above mean ****					
1. Q13718	Phytoene synthase.	1198	147	398	8.69	0
2. Q13359	Human mevalonate kinase gene.	1971	144	407	8.49	0
	**** 7 standard deviations above mean ****					
3. N60307	Equine IFN-beta from pAH60.	2467	135	349	7.88	0
4. Q43662	Acetobacter cdg3 operon.	4131	134	397	7.81	0
5. Q29265	Human calcium channel 27980/7	2050	132	388	7.68	0
6. N81083	Encodes human placental RNase	1698	131	407	7.61	0
7. Q05150	Modified human adipsin gene w	835	129	270	7.47	0
8. Q03566	Human adipsin/D cDNA hg 31-40	847	129	261	7.47	0
9. Q05149	Human adipsin gene from the c	1093	128	261	7.41	0
10. Q30966	Encodes vitamin K dependent c	2452	124	375	7.14	0

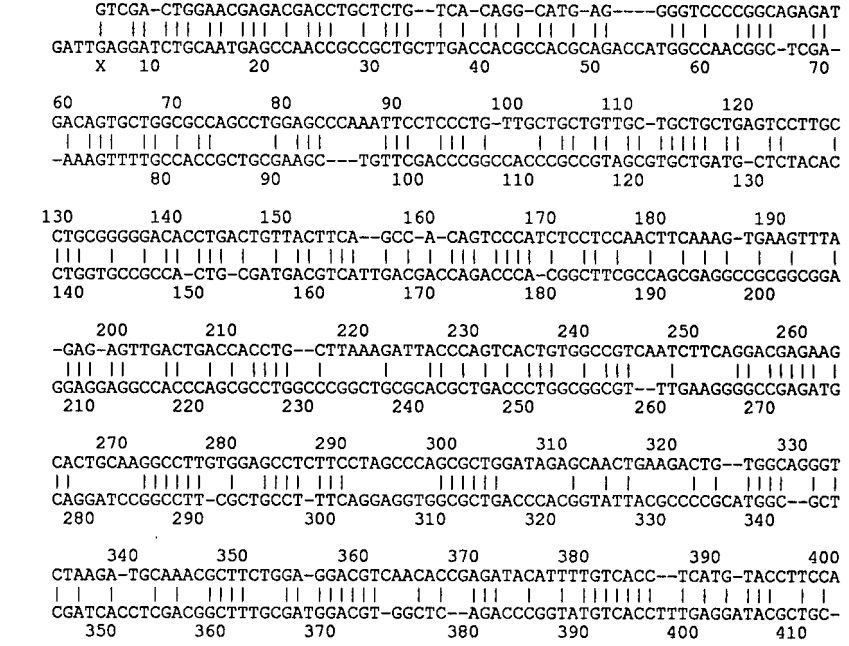
11.	Q25061	Clone 2918.4.	1102	123	382	7.07	0
12.	N70128	Novel DNA encoding a polypept	1363	123	387	7.07	0
13.	Q10613	Rianodin receptor gene.	15672	122	402	7.00	0
		**** 6 standard deviations above mean ****					
14.	Q37813	Sequence encoding the alpha 1	5904	120	396	6.87	0
15.	Q13798	Zea mays 15 kD seed storage p	1747	119	376	6.80	0
16.	Q23996	Mutant thermostable DNA polym	2043	117	344	6.67	0
17.	Q23994	Mutant thermostable DNA polym	2274	117	392	6.67	0
18.	Q23995	Mutant thermostable DNA polym	2364	117	392	6.67	0
19.	Q24223	Modified Taq polymerase codin	2499	117	392	6.67	0
20.	Q55406	A.chrysogenum beta-tubulin Ty	3445	116	409	6.60	0
21.	Q55405	A.chrysogenum beta-tubulin I1	3445	116	409	6.60	0
22.	Q48230	Acremonium chrysogenum beta-t	3445	116	409	6.60	0
23.	N81258	Encodes recombinant human pre	881	115	330	6.53	0
24.	Q53199	Human cyclin D3.	1963	115	416	6.53	0
25.	N81254	cDNA sequence in plasmid pT3B	2444	115	388	6.53	0
26.	Q11943	Nucleotide sequence of HIV-1	9192	115	363	6.53	0
27.	N82456	Sequence of mouse 1gG gamma 2	1946	114	392	6.46	0
28.	Q31875	Cyclin D3 gene.	1962	114	417	6.46	0
29.	N30041	Sequence encoding bovine leuk	1449	113	381	6.40	0
30.	Q58988	Human HSA-PDI fusion.	1545	112	388	6.33	0
31.	Q29619	HSA-PDI fusion protein.	1545	112	388	6.33	0
32.	Q58987	Human PDI.	2454	112	388	6.33	0
33.	Q29618	PDI.	2454	112	388	6.33	0
34.	Q14630	Plasmid pTM3 insert encoding	3131	112	397	6.33	0
35.	Q58776	Human glucagon receptor (pLJ6	1809	111	382	6.26	0
36.	Q05055	Placenta-specific protein PP1	2400	111	380	6.26	0
37.	Q05682	Islets of Langerhans cell clo	5051	111	389	6.26	0
38.	Q28399	Human nestin gene.	4854	110	388	6.20	0
39.	Q60714	Human brain Expressed Sequenc	397	109	156	6.13	0
40.	Q47354	3' end of cDNA clone of myoto	936	109	321	6.13	0
41.	Q45987	Prepro-HSA-G-CSF chimera enco	2382	109	388	6.13	0
42.	N90366	Synthetic human 5-lipoxygenas	2499	109	383	6.13	0
43.	Q06331	Human IL-2R beta chain.	4035	109	385	6.13	0
44.	Q05872	Sequence encoding human inter	4035	109	385	6.13	0
45.	Q04735	Gene fragment encoding human	1226	108	393	6.06	0

1. US-08-162-407-1 (1-879)
   Q13718 Phytoene synthase.

ID Q13718 standard; DNA; 1198 BP.
AC Q13718;
DT 26-NOV-1991 (first entry)
DE Phytoene synthase.
KW GGPP; carotenoid; phytoene; zeaxanthin; lycopene; ss.
OS Erwinia herbicola EHO-10 (E. vulneris - ATCC 39368).
FH Key Location/Qualifiers
FT CDS 16..939
FT /*tag= a
FT /label= phytoene_synthase
PN WO9113078-A.
PD 05-SEP-1991.
PF 04-AUG-1991; U01458.
PR 02-MAR-1990; WO-487613.
PR 18-MAY-1990; US-525551.
PR 03-AUG-1990; US-562674.
PA (STAD) AMOCO CORP.
PI Ausich RL, Brinkhaus FL, Mukharji I, Proffitt JH, Yarger JG;
PI Yen HC;

DR WPI; 91-281410/38.
DR P-PSDB; R13983.
PT Biosynthesis of carotenoid(s) in genetically engineered hosts -
PT using DNA encoding enzymes from Erwinia herbicola
PS Disclosure; Fig 4(1-3); 313pp; English.
CC There are a total of six relevant genes in a 7900 bp region that
CC cause E. coli cells to produce GGPP and the carotenoids phytoene
CC through zeaxanthin diglucoside, which is the final prod. identified
CC in the carotenoid pathway contd. in plasmid pARC376 (contg. a ca.
CC 13 kb chromosomal DNA fragment isolated by Perry et al., J. Bacteriol.,
CC 168:607 (1986). The genes for geranylgeranyl pyrophosphate (GGPP)
CC synthase, phytoene synthase, phytoene dehydrogenase-4H, lycopene
CC cyclase, beta-carotene hydroxylase, and zeaxanthin glycosylase are
CC represented in Q13716, Q13718, Q13719, Q13722, Q13724 and Q13726
CC respectively.
CC Recombinant expression plasmids can be used to produce large amts.
CC of the enzymes and hence large amts. of the carotenoids which they
CC synthesise.
SQ Sequence 1198 BP; 214 A; 371 C; 417 G; 196 T;

Initial Score = 147 Optimized Score = 398 Significance = 8.69
Residue Identity = 50% Matches = 472 Mismatches = 367
Gaps = 93 Conservative Substitutions = 0



```
          410       420       430       440       450       460       470
GCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCT
||  || | |  | ||  |||   |||   |      ||       |    ||| || |    |     |||
GCTACTGCTATCACGT-GGCG-GGCGTGGTGGGTCTGATGATGGCCAGGGTGATGGGCGTGCGGGATGAGC-
      420       430       440       450       460       470       480

      480       490         500       510       520       530
GCTTGCTCTGAAGC-CCTG-TATC--GGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTG-C
|  ||||     || ||||  |||  |||  ||||| ||||        |  |  || |||  |  | || |
GGGTGCTGGATCGCGCCTGCGATCTGGGGCTGGCCTTCCAGCTGACGAATATGGCCCGGGATAT-TATTGAC
      490       500       510       520       530       540       550

540       550       560         570       580       590       600
CAGCCGGACTCCTCCACCCTGCTGCC--CCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAG-
  |  ||| ||  |   | |||||  |  |||   ||||  |   ||    || |||  |  |  ||| |||
GATGCGG-CTATTGACCGCTGCTATCTGCCCGCCGAGTGGC--TGC--AGGATGCCGGGCTGGCCCCGGAGA
      560       570       580       590         600       610

610         620       630       640       650       660       670
CCTCGGC--CCAGGCAGCTGTTGCTCC-TGCTGCTGCTGCTGC-CTCTCACACTGGTGCTGCTGGCAGCCG-
  ||  ||  |  || ||    | || ||  ||||  || |  ||       |  | || ||| || |  |||||
ACTATGCCGCGCGGGAG-AATCGCCCCGCGCTGGCGCGGTGGCGGAGGCTTATTGATGCCGCAG--AGCCGT
620       630       640       650       660       670       680

   680       690       700         710       720       730       740
CCTGGGGC-CTTCGCTGGCAAAGGGCAAGA--AGGAGGGG-GGAGCTCCACCCTGGGGTGCCCCTCCCCTCC
  ||    | | || | |||   |||| | |   |   | || || ||||| |   || |  | || || | |
ACTACATCTCCTCCCAGGC--CGGGCTACACGATCTGCGGCGGCGCTCCGCGTGGGCGATCGCCACCGCCCG
690       700         710       720       730       740       750

      750         760       770       780       790       800
CATC-CCTA---GGA-TTCGAGCCTT--GTGCATCGTTGACTCAGCCAGGGTCTTATCTCGGTTACACCTGT
|| |  |||   |||  ||| |  ||  |   |  | |     || ||| | ||    |||    | || |
CAGCGTCTACCGGGAGATCG-GTATTAAGGTAAAAGCGGCGGGAGGCAGCGCCTGGGATCG---CCGCCAGC
760       770       780       790       800       810       820

810       820       830       840       850       860       870       X
AATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGAGCAGGTCGTCTCGTTC-CAGTCGAC
|  | ||||    | ||    |  ||      || ||   |  | || ||||||  | |||  |  || ||||
ACAC-CAGCAAAGGTGAAAAAATTGCCATGCTGATGGCGGCACCGGGGCAGGTTAT-TCGGGCGAAGACGAC
 830       840       850       860       870       880       890     X

GAGGGTGACG
 900
```

2. US-08-162-407-1 (1-879)
Q13359 Human mevalonate kinase gene.

```
ID   Q13359 standard; cDNA; 1971 BP.
AC   Q13359;
DT   12-NOV-1991  (first entry)
DE   Human mevalonate kinase gene.
KW   hMK; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             92..1191
FT   /*tag=  a
PN   GB2241500-A.
PD   04-SEP-1991.
```



```
PF   01-FEB-1991; 102244.
PR   06-FEB-1990; US-475577.
PA   (SQUI ) SQUIBB E R & SONS INC.
PI   Tanaka RD, Ricci BS, Mosley ST.
DR   WPI; 91-262010/36.
DR   P-PSDB; R13720.
PT   New nucleic acid encoding human mevalonate kinase - and derived
PT   polypeptide(s) for studying cholesterol biosynthesis, identifying
PT   cholesterol lowering agents or detecting mevalonate aciduria.
PS   Claim 1; Fig 1; 44pp; English.
CC   The DNA sequence was determined using the dideoxychain termination
CC   method. The sequence or homologous DNA can be labelled for use in
CC   detecting sequences encoding hMK, specifically for detecting the
CC   genetic disorder mevalonic aciduria which is characterised by low
CC   levels of hMK-coding sequences.
SQ   Sequence   1971 BP;    384 A;    601 C;    597 G;    389 T;

Initial Score    =    144  Optimized Score  =    407  Significance =  8.49
Residue Identity =    52%  Matches          =    501  Mismatches   =   317
Gaps             =    129  Conservative Substitutions              =     0

            X         10        20        30        40        50
            GTCGACTG--GAACGAGACGACCTGCTC-TGT-CACAGGCATGAGGGGTCCCCGGCAGAGAT
            || | |||  |   || ||   ||||||  || | |||  ||   ||  ||||  | ||  |
  CAGGGCCCTTGT-GGCTGGCGTCAGAAACAGGCTGCTCAAGTTCCCAGAGATCGTGGCCCCCCTCCTGACCT
820       830       840       850       860       870       880       890

  60        70        80        90              100       110       120
  GACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTG------TTGCTGCTGTTGCTGCTGCTGAGTCC
   | |  | || |  |  |||||||     |   | | |  |       | |  |  |   | ||  | ||| |
  CA-ATAGATGCCATCTCCCTGGAGTGTGAGCGCGTGCTGGGAGAGATGGGGGAAGCCCCAGCCCCGGAG-CA
           900       910       920       930       940       950       960

     130       140       150       160       170       180       190
  TTGCCT-GCG-GGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCT-CCTCCAACTTCAAAGTGAAGTT
   | ||| | |  || | | || | ||  || | | |  |||  ||||||    | |  | ||   |||  |
  GTACCTCGTGCTGGAAGAGCTCATTG--ACATGAAC--CAGCACCATCTGAATGC--CCTCGGCGTG--GGC
         970       980       990       1000      1010      1020

      200       210       220       230       240       250       260
  TAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCA-ATCTTCAGGAC--GAGAA
   |  |  |     ||||| || ||   || | | ||||   | | || | ||| | |      |||  | | |
  CA-CGCCTCTCTGGACCAGCT-CTGCCAGGTGA-CCAGGGCCCGCGGACTTCACAGCAAGCTGACTGGCGCA
     1030      1040      1050      1060      1070      1080      1090

        270       280       290       300       310       320
  -GCACTGCAAGGCCTTGTGG-AGC-CTCTTCCTAGCCCAGCGCTGGA-TAG--AGCAACTGAAGAC--TGTG
   ||  ||   |||  ||||| | | | || |  |  |||| ||||||  ||  || || || || |   |
  GGCGGTG-GTGGC--TGTGGCATCACACTCCTCAAGCCAGGGCTGGAGCAGCCAG-AAGTGGAGGCCACGAA
     1100      1110      1120      1130      1140      1150      1160

   330       340       350       360       370       380       390
  GCAGGGTCTAAGATGCAAACGCTT---CTG-GAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTAC
  |||||  || |   ||    ||||   |||   |||   || || ||      |     ||| |  |||  ||
  GCAGGCCCTGACCAGCTGTGGCTTTGACTGCTTGGAAACCAGCATCGGTGCCCCCGGCGTCTC--CATCCAC
       1170      1180      1190      1200      1210      1220      1230

     400       410       420       430       440       450       460
  CTTCCAGCC-CCTACCAGAATGTCTGCGATTCGTCCAGACCAA--CATCTCCCACCTCCTGA-AGGACACCT
    || |||| ||| || | |   |  |||   ||||||  |||  | |      ||| |||| ||||  ||
```

```
    --TC-AGCCACCTCCCTGGACAGC--CGA---GTCCAG--CAAGCCCTGGATGGCCT-CTGAGAGGAGCCCA
           1240      1250           1260        1270        1280      1290

         470        480       490         500         510        520
    GCACACAGCTG-CTTGCTCTGAAGCCCTGTATCGGGA--AGGC--CTGCCAGAATTTC-TCTCGGTGCCTGG
      |||| || | |   | | || ||||  | || |||  |  |    | | | | |||  ||| ||| ||||
    CGACACTGCAGCCCCACCCAGATGCCC-CTTTCTGGATTATTCTGGGGGCTGCAGTTCGACTCTGTG-CTGG
        1300      1310       1320       1330       1340       1350       1360

   530       540       550       560       570       580       590
    AGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCC-CTAGAAGCCACGGAG
         ||| |   |||   ||||| || ||||||      || || |||  |||| || |  || | |
    ----CCAGCGAGCGCCCAGCTCCT-GACACTGCTG-----GAG-AGGCCC-CAGCCGCTTG--GCGATG--C
           1370      1380       1390            1400       1410

  600       610       620       630       640           650           660
    CTCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGC-TG--CTG-CCTCT-CA--CACTGGTG-C
    |  ||| || ||| ||     |||| || |  |||   |  || ||  ||| ||||| ||  | |||| | |
    CAGCCA-AG-CTCTGCAGTCCCAGCGGTGGGACCTAGGGAGGCATGGTCTGCCCTCTGCATCCTCTGGAGCC
  1420        1430      1440      1450      1460      1470      1480

        670       680       690       700       710           720
    TGCTG-GCA-GCCGCCT-GGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGA---GCT-CCACCCTGGG
     || | ||| |  |||| || |  | | ||||  | |   ||      |||| | ||   ||| |  ||||  |
    AGCCGAGCAGGAGGCCTAGGAGGGTCCTCTG--AGACTCCAGACCTGAGGCGAGAAGGGCTGCTTCCCT--G
  1490      1500      1510      1520      1530      1540      1550

   730       740         750       760       770           780       790
    GTGCCCCTCCCCTCCCATC--CCTAGGATTCGAGCCTTGTGCATCGT--TGACT--CAGCCAGGGTCTTAT-
      || ||  |  |||||||  | |  |   |  ||||||  |   ||  |  ||  | ||| |||||  ||
    AAGCTCCCACAGTCCCATCTGCTTCAGGCCCCCGCCTTGGCCTGTGTTCTTCCTGGCCGCCTGGGTCCAATG
    1560      1570      1580      1590      1600      1610      1620

       800       810       820       830       840       850       860
    CTCGGTTACACCTGTAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGAGCAGGTC---
    ||| | |    |||    ||    ||   ||||    | ||    |  ||    |  || ||  | || |
    CTCAGGT---GCTGGGGCCTGGTTCC--CGGAGAAGTGTGCCTTCTCTCT-CCTTTTCAGGGACCGGCCCCC
  1630        1640      1650        1660      1670      1680      1690

           870       X
    -GTCTC---GTTCCAGTCGAC
     |||||   |  ||||  |  |
    TGTCTCTCAGGGCCAGGCCTCTCCCTCCTCC
      1700      1710    X 1720


3. US-08-162-407-1 (1-879)
   N60307       Equine IFN-beta from pAH60.

 ID   N60307 standard; DNA; 2467 BP.
 AC   N60307;
 DT   10-JUN-1991  (first entry)
 DE   Equine IFN-beta from pAH60.
 KW   IFN-beta; equine; interferon; ss.
 OS   Equus cabalus.
 FH   Key             Location/Qualifiers
 FT   CDS             677..1254
 FT   /*tag=  a
 FT   /product= IFN-beta
 FT   sig_peptide     697..759
```



```
 FT   /*tag=  b
 FT   mat_peptide     760..1254
 FT   /*tag=  c
 FT   polyA_signal    1446..1451
 FT   /*tag=  d
 FT   polyA_signal    1516..1521
 FT   /*tag=  e
 FT   polyA_signal    1705..1710
 FT   /*tag=  f
 FT   misc_RNA        2441..2467
 FT   /*tag=  g
 FT   /note= "pUC9 start"
 PN   EP-186098-A.
 PD   02-JUL-1986.
 PF   17-DEC-1985; 116083.
 PR   18-DEC-1984; DE-446122.
 PA   (BOEH ) BOEHRINGER INGELHEIM.
 PI   Himmler A, Hauptmann R, Hauel N, Adolf G, Swetley P.
 DR   WPI; 86-170649/27.
 DR   P-PSDB; P60398.
 PT   New equine and canine interferon - and recombinant DNA molecules
 PT   coding for them, and transformed cells.
 PS   Disclosure; Fig 8; 149pp; German.
 CC   Microorganisms transformed with the recombinant sequence produce IFN
 CC   which is useful therapeutically in veterinary medicine.
 CC   See also N60306-13 and N60938.
 SQ   Sequence   2467 BP;    720 A;    563 C;    556 G;    628 T;

Initial Score    =    135  Optimized Score  =    349  Significance =  7.88
Residue Identity =    49%  Matches          =    418  Mismatches   =   320
Gaps             =    103  Conservative Substitutions               =     0

             X       10              20        30        40        50
             GTCGACTGGAAC------GAGAC-GACCTGCTCTGTCACAGGCATG-AGGGGTCCCCGGCAG
             || | || || |      | || ||  |   ||  | || || | || ||  |  | |
   GAGAAGAAATGT-GGCTTGAGCCCCATATAAACGGAATTAAAATGGGAGAGACAGGCAGAGGCTCTGGACTC
1670      1680       1690      1700      1710      1720      1730      1740

       60          70        80        90        100       110       120
   AGATGACAGTGCTG--GCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTT-GCTGCTGCTGAGTC
   ||| ||| | ||||  ||  | |||  | | ||      ||| |||   |  | ||| |   ||| ||  ||
   AGAGGACGGGGCTGCTGCTTCTGCCCTGTGTCC----CGCTCTCTGGCCCCACAGTTAGAATCTGATGGCTC
        1750      1760      1770          1780      1790      1800

      130         140       150       160       170       180       190
   CTTGCCTGC---GGGGGACACCTGA-CTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAG
     |  |||   | || | |  | | || || | |  |||     | |||| ||| | |  ||   | || |
   TCAGGGTGCCCAGAGGAATATGTCAGCTCTTGCGTTTGCCTGGAGCTCATC-CCTACTA--TCTGCGAGATG
 1810      1820      1830      1840      1850      1860      1870

        200       210       220       230       240       250       260
   TTTAGAGAGTTGACTGACCACCTGCTTA-AAGAT--TACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAG
    |  |          |  || |   |  | | |||  ||      |  ||||| ||  ||| | | ||   |||
   CTCTGCCTCCCCCCACCCCTCAACCCCACAGGATTGTAAAATATTTCTGTGCCCTGCAAGCCTAAGCGGGAG
 1880      1890      1900      1910      1920      1930      1940

        270         280          290          300          310
   AAGCACTGCAAGGC-CTTGT-GGAGCCTCT---TCCTAGCCC----AGCG----CTGGATAGAGCAACTGAA
   |||   | | |||| ||| | |||  || | |   || ||    |  |    ||   | |  |||| ||
   AAG---TCCCAGGCACTTCTGGGACACTGTAAGTGGCAGTCCCTTTATGGTACTCTTCTTGGGACAACCGA-
```

```
1950          1960       1970       1980       1990       2000       2010

 320       330          340        350        360          370        380
   GACTGTGGCAGG-GTCTAA--GATGCAAACGCTTCTGGAGGACGTCAA---CACCGAGA-TACATTTTGTCA
     ||||  |||| ||||||  |  || | ||  ||| |    | || |   ||| || |  | |    |  |
   -GCTGT-ACAGGTGTCTAAGGGGAGCCAGCGTCTCT-GTCTCCTTCCAGGGCACAGACACAAGAGGAAGAAA
   2020        2030       2040       2050       2060       2070       2080

        390        400        410        420        430        440        450
   --CCTCATGTACCTTCCAGCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAA
      ||| |||    |||   ||||| |||    | |||   ||  ||    || |  | |||         |||
   GAACTC-TGT---TTCATACCCCTGCCA-TCGGCCTGGTTTTGCTC----CCTATTTTTCC-----AGAGAA
   2090           2100       2110       2120           2130           2140

     460        470        480        490        500        510        520
   GGACACCTGCACACAGCTGCTTGCTCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCT
   | |   |||| | | |||  || | |||  || |||  ||    ||||    || | | || | |   |||
   GCAAGTCTGCTC-CGGCT--TTTC-CTGCTGCTCTGCGTC-TCCAGGC----CACACTCTC-CCCAAAGCCA
     2150       2160          2170       2180           2190       2200

      530        540        550        560        570        580        590
   GG---AGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCAC
     |   | | ||||   |||| |     ||||  || | | ||    ||  ||    |  | | | || ||
   AGGCCAAG-GCAGGTGCAGCAGATTAGGTCCAGTCT-CAGGCCAGTGAGAAACCGGGAAGCATGGGAGACAA
   2210       2220       2230       2240       2250       2260       2270

     600        610        620        630        640        650        660
   GGAGCTCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACTGGTG--CTG
   |||  | |  | |     |    |||| ||   | |||    |||  | |   ||   | | ||| |   |
   GGAAATTCAGGTGGGATAGAGAGGGCA-CTAACGTTCCCAGGGCT--TAC--ACTGGGAAATTGGAGATTTC
   2280       2290       2300       2310       2320          2330       2340

     670        680        690        700        710        720        730
   CTGGCAGCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGGGGTGCCCCTC
   || | || | | | ||||    ||  ||| |  ||          |   | |||  ||||| ||  |  | |
   CTAGGAG-CTCTTTGGGCACCGGCCAGCATA--GCTGCTTTTCTGTCTGTGCTGAACCCTTGG--GAACGTG
      2350       2360       2370       2380       2390       2400       2410

   740        750          760        770          780        790        800
   CCCTCCCATCCCTAGGATT-CGA-GCCTTGTGCAT-CGTTGAC--TCAGCCAGGGTCTTATCTCGGTTA
   |  |   || |||    || | |  |    | | || ||| |||   |||||| |  |||
   CATTATTATGCCTGTTCTTGCCATGAGCAGGGGATCCGTCGACCTGCAGCCAAG--CTT
         2420       2430       2440       2450       2460        X
```

4. US-08-162-407-1 (1-879)
   Q43662 Acetobacter cdg3 operon.

```
ID   Q43662 standard; DNA; 4131 BP.
AC   Q43662;
DT   13-OCT-1993  (first entry)
DE   Acetobacter cdg3 operon.
KW   Cyclic diguanylate; diguanylate phosphodiesterase;
KW   diguanylate cyclase; cellulose production; ss.
OS   Acetobacter xylinum.
FH   Key             Location/Qualifiers
FT   CDS             387..2606
FT   /*tag=  a
FT   /label= pdeA3
FT   /note="a diguanylate phosphodiesterase gene;
```



```
FT   the amino acid sequence R38155 cannot be
FT   deduced directly from this open reading
FT   frame due to a number of apparent
FT   (typographical ?) deletions in the printed
FT   sequence"
FT   misc_difference 741..742
FT   /*tag=  b
FT   /codon= seq: GT; aa: Val
FT   /note= "insertion of N between nucleotides 742 and
FT   743 will restore the reading frame"
FT   misc_difference 2522..2523
FT   /*tag=  c
FT   /codon= seq:GG; aa: Trp
FT   /note= "insertion of T between nucleotides 2521 and
FT   2522 will restore the reading frame"
FT   misc_difference 2560..2561
FT   /*tag=  d
FT   /codon= seq:CC; aa:Ala
FT   /note= "insertion of G between nucleotides 2559 and
FT   2560 will restore the reading frame"
FT   CDS             2657..4131
FT   /*tag=  e
FT   /label= dgc3
FT   /note= "a diguanylate cyclase gene;
FT   the ORF is described as continuing to
FT   nucleotide 4143 although the sequence is
FT   only shown up to nucleotide 4231. The amino
FT   acid sequence R38156 cannot be deduced
FT   directly from this open reading frame which
FT   is given as printed in the specification"
FT   misc_difference 3848..3850
FT   /*tag=  f
FT   /codon= seq:GGG; aa:Ala
FT   misc_difference 3962..3963
FT   /*tag=  i
FT   /codon= seq:CC; aa: Thr-Phe
PN   WO9311244-A.
PD   10-JUN-1993.
PF   14-OCT-1992; U08756.
PR   29-NOV-1991; US-800218.
PA   (WEYE ) WEYERHAEUSER CO.
PI   Ben-Bassat A,  Benziman M,  Calhoon RD,  Gelfand DH;
PI   Tal R,  Wong HC;
DR   WPI; 93-197062/24.
DR   P-PSDB; R38155, R38156.
PT   Polynucleotide sequence from Acetobacter cdg operon - encodes
PT   cyclic di:guanosine mono:phosphate degradation enzymes e.g.
PT   3-phosphodiesterase isozyme
PS   Claim 3; Page 70-72; 98pp; English.
CC   A gene bank of Acetobacter 1306-11 was constructed in the broad host
CC   range, mobilisable cosmid pKT230C0S5. The average insert size was ca.
CC   30kb. The bank was screened with a 53bp PCR amplified probe whose
CC   sequence was expected to be an exact match for a diguanylate
CC   cyclase encoding gene. Three unique cosmid classes were identified
CC   corresponding to three different cdg operons. The cdg3 operon
CC   comprises at least 2 genes. Manipulating the expression levels of
CC   the different cdg3 genes is contemplated, particularly in cellulose
CC   producing bacteria.
SQ   Sequence   4131 BP;    864 A;    1275 C;    1197 G;    795 T;
```

```
Initial Score     =    134  Optimized Score  =    397  Significance =  7.81
Residue Identity  =    52%  Matches          =    486  Mismatches   =   323
Gaps              =    124  Conservative Substitutions               =     0

             X       10          20          30          40          50
             GTCGACTGGAACGAGACG--ACCTGCTCTGTCA-CAGGC-ATGAGGGGTCCCCGGCAGAGAT
              |   || |||  | | |  | |||| | ||| ||  | ||||     |  |||| |  ||
   TGGTATAATGATTATCTTGAATAATATGTAATATGCT-TATCATCATCCTATGATATTCCTGCGGCCGCCAT
      210    X  220         230         240         250         260         270

       60          70          80          90         100         110         120
      G-ACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGTCCTTGCC
      | || ||||   |||| |  ||  | |       | |||||    | | || ||  ||  | | ||   ||
      GCACTGTGC--CCGCCCGGGTGATGGC------GCGCCCTGCATTTCCGGTCGC--CTTAT-ATACC---CC
        280         290         300         310         320         330

   130         140         150         160         170         180         190
      TGCGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCC--TCCAACTTCAAAGT-GAAGTTTAGAG
      ||| |  | | |||| | ||  || | ||   |||| | | |     ||| |||| ||  |||  ||
      TGCTGCAGCCGCCTGCCGGTGGCTGC--CCCTTGTCCCTTTTGCGAGGTAAC-TCAATGTCAAAGCCTAACC
          340         350         360         370         380         390         400

    200         210         220         230         240         250         260
      AGTTGACTGACCACCTGCTTAAAG-ATT--ACCCAGTCACTGTGG---CCGTCAATCTTCAGGACGAGAAGC
       |    ||||   |||| ||   | | |  | |  | ||  |  |    |||| | | || | ||| ||||
      TGAGCGCTGAAGTCCTGATTTCGGCACTGGAGCAGGCCATAGACGCAACCGTGATTATTGATGACAAGAACG
           410         420         430         440         450         460         470

       270         280         290         300         310         320         330
      ACTGCAAGGCCTT--GTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTC
      |   ||    |||   |   ||    ||  |  || | |  |  | || |||   |||| ||| || |||  |
      AAGTCATCTTCTTCAATCAGGCGGGATCGAA--CCTGTGGGGCCTCGA-CAA-GAAAGA-TGT-GCTGGG-C
         480         490         500         510         520         530

       340         350         360         370         380         390
      TAAGATGCAAACGCTTCTGGAG--GACG--TCAACA--CCGAG--ATAC-ATTTTGTC-ACCTCATGTACCT
       |  | |  |||| | |||| |   |||  || ||  ||| |    ||   ||||||  | | ||    |
      AAGAACGTGAACG-TGCTGGTGCCCACGCTGCATCAGGCCGGGCACGACGCCTTTGTCGAGCGCA--GCCGG
    540         550         560         570         580         590         600

       400         410         420         430         440         450
      TCCAG--CCCCTACCAG-AATGTCTGCGATTCG-TCCAGACCAACATCTCCCACCTCCTGAAGG-----ACA
        |||  | | || | | | |||| ||   |||  | ||    |  || |||  |||  || ||      | |
      GGCAGCTCGCATAACCGCATTGTCGGCACCTCGCGCGAGGTGGAGTTCACCC-GCTC-GGATGGGGAATATA
    610         620         630         640         650         660         670

    460         470         480         490         500         510         520
      CCTGCACACAGCTGCTTGCTCT-GAAGCCCTGTATCGGGAAGGCCTGC-CAGAATTTCTCTCGGTGCCTGGA
       ||||    ||||| | ||| |  |||  |   | |  || | |  || ||   |||   | || |  || |
      TCTGCGGCGAGCTG-TCGCTGTCCAAGGTCGTGAAC--GATGACAAGCGCATCTTTTTCATGGGCGTATGAA
    680         690         700         710         720         730         740

    530         540         550         560         570         580         590
      GGTGCAGTGCCAGC--CGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCAC-GGA
      |    | || || |  || |  ||  || |  ||      ||   || |||   ||  | |   ||  | ||
      G----AATGTCACCAACGAAAGCCAGCAGC--GCAAGATCC--TGA-TCCTGCAGAACGACGTGCTGCAGGC
          750         760         770         780         790         800         810

    600         610         620         630         640         650         660
      GCTC-CCAGAGCC---TCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACAC-TGGTGCTG
```



```
       ||| |||| | |     |  ||||| |  ||     ||| ||| ||| || |  |   |  | | ||||
      ACTCGCCAGCGACATGATGATCCAGG-ACGTGGCCGACCTTCTG-TGCCGCCGGGTGGAAAGCTTCGTGC--
            820         830         840         850         860         870

        670         680         690         700         710         720         730
      CTGGCA-GCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGGGGTGCCCCT
      | |||| |  | | | |  | | |||  |     || |||       ||| || | | |   |  || ||
      CCGGCACGGTGGCCGTGCTCAT-GCTG--ATTACGCCAGA-----CGGGCAGTTGCGC---GTTCTGTCCAG
    880         890         900         910         920         930

        740         750         760         770         780         790         800
      CCCCTCCCATCCCTAGGATTCGAGCCTTGTGCATCG-TTGACTCAGCCAG--GGTCTTATC-TCGGTTACAC
      |||| | |  ||| ||     ||   |  | || ||| | ||   |||| |   ||||  || || |  || |
      CCCCACACTGCCCAAG----CGCTAC-CGCGCCTCGCTCGA--GAGCCTGTATGTCTCGTCATCCG-AACTC
  940         950         960         970         980         990        1000

          810         820         830         840         850         860         870
      CTGTAATCTCAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGAGCAGGT-CGTCTCGTTC--CA
        | || ||  || | |  || |||| |||| |    | | ||||| |||  ||| | |  |||| ||  ||
      GAG-AAGCT--GCGCGT-CGATCCCA-AGCACGCCACCCGCATGGTGTGGGACAGCTACCGCTCGCTCGGCA
         1010        1020        1030        1040        1050        1060        1070

          X
      -GTCGAC
        |||
      TCTCGCTTGGCCTGCAG
           1080
```

5. US-08-162-407-1 (1-879)
   Q29265 Human calcium channel 27980/7.

```
ID   Q29265 standard; DNA; 2050 BP.
AC   Q29265;
DT   03-MAR-1993  (first entry)
DE   Human calcium channel 27980/7.
KW   Plasmid pR9112-10.1.1.1; Ca-flux assay; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_difference 1463
FT   /*tag=  a
FT   /note= "undefined"
FT   misc_difference 1940
FT   /*tag=  b
FT   /note= "undefined"
FT   misc_difference 1941
FT   /*tag=  c
FT   /note= "undefined"
FT   misc_difference 1942
FT   /*tag=  d
FT   /note= "undefined"
PN   EP-507170-A.
PD   07-OCT-1992.
PF   23-MAR-1992; 104970.
PR   04-APR-1991; DE-110785.
PA   (FARB ) BAYER AG.
PI   Franz J, Rae P, Unterbeck A, Weingaertner B;
DR   WPI; 92-333446/41.
DR   P-PSDB; R27645.
PT   Cloned human neuronal calcium channel sub-types - useful in
```

```
PT   calcium flux assays to screen for neurone-specific calcium
PT   channel ligands
PS   Claim 2; Page 45-49; 101pp; German.
CC   Human neuroblastoma cell line, hippocampus, frontal and temporal
CC   cortex and visual cortex cDNA banks were screened with a probe
CC   containing carp skeletal muscle Ca-channel cDNA. The cDNA clone
CC   pR9112-10.1.1.1 was sequenced and found to encode amino acids 991 to
CC   1650 containing part of the Domain III (s6), the whole fourth
CC   domain and a part of the C-terminal cytoplasmic region of the
CC   protein. This clone differs from pR9112-2.1.1.1 (see Q29267) and
CC   pR9112-4.1.1.1 (see Q29264) by a 57bp insert with splicing
CC   consensus sequences at both ends, thus offering the possibility of
CC   alternative splicing. The sequence can be inserted into a eukaryotic
CC   expression vector for use in transforming suitable host cells. Cell
CC   lines producing human neuronal calcium channel proteins can be used
CC   for screening for Ca channel ligands (agonists or antagonists).
CC   See also Q29259-Q29275.
SQ   Sequence   2050 BP;    464 A;    617 C;    532 G;    433 T;
SQ   4 Others;

Initial Score     =    132  Optimized Score  =    388  Significance =  7.68
Residue Identity  =    50%  Matches          =    473  Mismatches   =   338
Gaps              =    126  Conservative Substitutions              =     0

      X    10        20        30        40        50        60        70
   GTCGACTGGAACGAGACGACCTGCTCTGTCACAGGCATGAGGGGTCCCCGGCAGAGATGACAGTGCTGGCGC
      ||   |||| || | |  |   || | |||   ||    | ||   |||||  |||| |  ||
      GA---GAAC-AG-CAAGTTTGACTTTGACA---AT----GTTC--TGGCAGCCATGATGGCCCTCTTCA
      X           10        20                  30        40        50

          80        90       100       110       120       130       140
   CAGCCTGGAGC-CCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGC--TGAGTCCTTGCCTGCGGGGGACAC
   | | ||  | |  | ||       || |  ||||||| | | ||| |   || |||   |  |||  ||||
   CCGTCTCCACCTTCGAAGGGTGGCCAG-AGCTGCTG-TACCGCTCCATCGACTCC---CACACGGAAGACAA
      60        70        80        90       100       110       120

         150       160       170       180       190       200       210
   CTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAGTTTAGAGAGTTGACTGACCACC
     | |.  | || || | || ||        |||||  ||||    | |  | || |   |    | | | ||
   GGGCCCCAT-CTACAACTACCGTGTGGAGATCTCCATCTTCTTCATCA--TCTACATCAT--CATCATCGCC
        130       140       150       160       170       180

      220       230       240                 250               260       270
   TGCTTAAAGATTACCCAGTCACTGTGGC--CGTCA--ATC----TTCAGGA------CGAGAAGCACTGCAA
   | ||| | ||| | |   |  || |||  |||||   ||    |||||||       ||| || | | |||
   TTCTTCATGATGAACATCTTCGTG-GGCTTCGTCATCGTCACCTTTCAGGAGCAGGGGGAGCAGGAGTACAA
190       200       210       220       230       240       250       260

             280       290       300       310       320       330
   GGCCT-TG---TGGAGCCTCTTCCTAGC-CCAGCGC-TGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAA
   |  || ||   |||| |     || |||  ||| || |||| | |    || ||| |   |||     |  |
   GAACTGTGAGCTGGA-CAAGAACC-AGCGACAGTGCGTGGA-ATACGCCCTCAAGGC-CCGGCCCCTGCGGA
          270       280       290       300       310       320

   340       350       360       370       380       390       400
   GATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCTACC
   | | ||     | |   || ||  || ||||  | | | | ||  || |  |  ||  |||| |  || |
   GGTACA-----TCCCCAAGAAC--CAGCACCAGTACAAAGTGTGGTACGTGGTCAAC--TCCACCTACTTCG
 330             340       350       360       370       380       390
```



```
410        420       430       440       450       460       470
  AGAA--TG-TCTGCG-ATTCGTCCAGACCAACATCTCCCACCT--CC-TGAAGGACACCTGCACACAGCTGC
  || |  || | | ||   || ||| |  ||||| |  |  |||  || || || ||  | || || ||||||
  AGTACCTGATGTTCGTCCTCATCCTGCTCAACACCATCTGCCTGGCCATGCAGCACTACGGC-CAGAGCTGC
       400       410       420       430       440       450       460

     480       490       500       510       520       530       540
  TTGCTCTGAA--GCCCTG--TATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAG-TGCCA
   || ||  ||  ||| ||   |||   ||  | |||       ||| || |   | |    ||| || ||  |
  CTGTTCAAAATCGCCATGAACATCCTCAA--CATGC----TCTTCACTGGCCTCTTCACCGTGGAGATG--A
      470       480       490               500       510       520

           550       560       570       580       590       600       610
  GCCGGA--CTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCT
   || ||  ||| |    ||| |    ||||||  |   | ||  |  | || | || |||  | |    | |
  TCCTGAAGCTCAT--TGCCTTC-AAACCCAAGCA---CTATTTCTGT-GATG-CATGGA-ATACATTTGAC-
   530         540       550       560       570       580

         620       630       640       650       660       670
  CGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACTGGT----GCTGCTG-GCAGCCGCCT
    |||  |   | |||  | |     || || |  |  |  ||||   |||      | ||||  ||  | |
  --GCCTTGATTGTTGTGGGTAGCATTGTTGAT-ATAGCAATCACCGAGGTAAACCCAGCTGAACATACCCAA
   590       600       610       620       630       640       650

 680       690       700       710       720       730       740
  GGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGAGCTC---CACCCTGGGGTGCCCCTCCCCTCCCATC
   |  |   | ||   || |   | |||        | | |||   |||| |     |||  | ||    |||
  TGCTCTCCCTCTATGAACGCAGAGGAAAACTCCCGCATCTCCATCACCTTCTTCCGCCTGTTCCGGGTCATG
  660       670       680       690       700       710       720

    750           760           770       780       790       800
  C--CTAG-GATTC----GAGCCTT-----GTGCATCGTTGACTCAGCCAGGGTCTTATCTCGGTTACACCTG
  |  || | ||  |    |||||  |     | |||||   ||| | ||   || | |   ||   |  | |
  CGTCTGGTGAAGCTGCTGAGCCGTGGGGAGGGCATC-CGGACGCTGCTGTGGACCTTCATCAAGT--CCTTC
730       740       750       760       770       780       790

  810       820       830       840       850       860       870
  TAATCTCAGCCCT-TG-GGAGC-CCAGAGCAGGATTGCTGAATGGTC-TGGAGCAGGTCGTCTCGTTCCAGT
   |  | | ||||| || ||  | || || |  || ||||||  |  || | |  |||  ||  | | |||
  CAGGCCCTGCCCTATGTGGCCCTCCTGATCGTGA-TGCTG-TTCTTCATCTACGCGGTGATCGGGATGCAGG
800       810       820       830       840       850       860

    X
  CGAC
   |
  TGTTTGGGAAAATT
870 X     880
```

6. US-08-162-407-1 (1-879)
   N81083 Encodes human placental RNase inhibitor

```
ID   N81083 standard; DNA; 1698 BP.
AC   N81083;
DT   30-OCT-1990  (first entry)
DE   Encodes human placental RNase inhibitor
KW   angiogenin inhibitory activity; neovascularisation disorders;
KW   human placental RNase inhibitor (PRI); rheumatoid arthritis;
KW   Kaposi's sarcoma; ss.
OS   synthetic.
```

```
FH   Key             Location/Qualifiers
FT   CDS             122..1507
FT   /*tag=  a
FT   /product=human placental RNase inhibitor
PN   EP-291686-A.
PD   23-NOV-1988.
PF   12-APR-1988; 105781.
PR   14-APR-1987; US-038008.
PR   5-APR-1988; US-177942.
PA   (HARD) Harvard College.
PI   Shapiro R, Vallee BL;
DR   WPI; 88-331302/47.
DR   P-PSDB; P80646.
PT   Inhibitors of angiogenin, useful for inhibiting tumour growth -
PT   include human placental RNase inhibitor and active polypeptide
PT   segments
PS   Disclosure; p; English.
CC   cDNA sequence isolated from human placental library using probes
CC   N81071 to N81082. Tryptic digestion of the PRI protein encoded by
CC   this sequence produces peptide fragments possessing angiogenin
CC   inhibitory activity.
CC   Disorders associated with neovascularisation such as rheumatoid
CC   arthritis and Kaposi's sarcoma are treated by admin of these
CC   inhibitory peptides.
SQ   Sequence  1698 BP;   316 A;   535 C;   535 G;   312 T;

Initial Score    =    131  Optimized Score  =    407  Significance =  7.61
Residue Identity =    52%  Matches          =    493  Mismatches   =   328
Gaps             =    126  Conservative Substitutions              =     0

            X        10        20        30        40        50
            GTCGAC-TGGAACGAGACGACCTGCTCTGTCACAGGCATG---AGGGGTCC-CCGGCAGAGA
             | ||| |  |  | || || |||  | | | |  | |||    | | ||| ||  | |
    CATCCAGAGCCTGGACATCCAGTGTGAGGAGCTGAGC-GACGCTAG-ATGGGCCGAGCTCCTCCCTCTGCTC
         140  X    150       160       170       180       190       200

     60        70        80        90       100       110       120
    TGACAGTGC--TGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGTCCTTG
       ||||||   |  | ||| |||||   |     |||| | |  ||  | ||     |    | ||  ||
    CAGCAGTGCCAAGTGGTCAGGCTGGACGACTGTGGCCTCACGGAAGC-ACGGTGCAAGGACATCAGCTCTGC
         210       220       230       240       250       260       270

    130       140       150       160       170       180       190
    CCTGCGGGGGACACCTG-ACTG--TTACTTCAGCCACAGTCCCATC-TCCT--CCAACTTCAAAGTGAAGT-
     || || |  |  |||| ||||    |  ||| ||   |   || |   ||   | |  ||   ||| | |
    ACTTCGAGTCAACCCTGCACTGGCAGAGCTCAACCTGCGCAGCAACGAGCTGGGCGATGTCGGCGTGCATTG
        280       290       300       310       320       330       340

             200       210       220       230       240       250
    ----TTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCA-ATCTTCAG---GA
         |  || |    | |||| ||| ||  |||||   ||||   |||  |||  || | || | |   ||
    CGTGCTCCAGGGCCTGCAGACCCCCTCCTGCAAGAT---CCAGAAGCTG-AGCCTCCAGAACTGCTGCCTGA
      350       360       370       380           390       400       410

     260       270       280       290       300       310       320
    CGAGAAGC-ACTGCAAGGCCTTGTGGAGC-CTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTG
    || |   |  ||||  || | |||  ||| | || |  | ||| || |   |    |||   | || ||
    CGGGGGCCGGCTGCGGGGTCCTGTCCAGCACACTACGCA-CCCTGCCC---ACCCTGCA---GGAG-CT---
        420       430       440       450       460             470
```



```
      330       340       350       360       370       380       390
    GCAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCA---CCTCATGTACC
    |||    || ||   ||| | ||  ||| ||| |   |    ||  | |  | ||  |   ||| || |||
    GCA---CCTCAGCGACAACCTCT--TGGGGGATGCGGGCCTGCAGCTGC--TCTGCGAAGGACTCCTGGACC
          480       490       500       510       520       530

      400       410       420       430       440       450       460
    TTCCAGCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTC--CCACCTCCTGAAGGACACCTGCA
      ||||  ||  |  |||   ||||   | |   |  || | ||||  |  ||  |||  |  | ||||
    -CCCAGTGCCGCCTGGAAAAGCTGCAGCTGGAGTATTGCAGCCTCTCGGCTGCCAGCTG-CGAGCCCCTG-G
  540       550       560       570       580       590       600

      470       480       490       500       510       520       530
    CACAGCTGCTTGCTCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGT--GCCTGGAGG--TG
    |    || | ||||| |  |||    | | |||    || | || |   ||| ||||  ||    | |
    C----CTCCGTGCTCAG-GGCC---AAGCCGGA----CTTCAAGGA--GCTCACGGTTAGCAACAACGACAT
       610       620       630       640       650       660

        540       550       560       570       580       590       600
    CAGTGCCAGCCGGACTCC-TCCACCCTGC--TGCCCCCAAGGAGTCCCATAGCC--CTAGAAGCCACGGAGC
    || ||   || ||  ||| |   |  |||   | ||  ||||| ||||   |||  || || ||    |||
    CAATG-AGGCTGGCGTCCGTGTGCTGTGCCAGGGCCTGAAGGACTCCCCCTGCCAGCTGGAGGCGCTCAAGC
     670       680       690       700       710       720       730

           610       620       630       640       650       660
    TCCCAGAGCCTCGG------CCCAGGCAGCTG-TTGCTCCTGCTGCTGC--TGCTGCCTCTCACACTGGTGC
    |   |||||  |||        |||  || |||   |  |||| ||| ||  ||  ||||| ||       ||
    T-GGAGAGCTGCGGTGTGACATCAGACAACTGCCCGGGACCTG-TGCGGCATTGTGGCCTC-CAAGGCCTCGC
     740       750       760       770       780       790       800

        670       680       690       700       710       720       730
    TGCTGGCAGC-CGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGG-AGGGGGGAGCTCCACCCTGGGGTGCC
    ||| || |||  |||  ||||     |  ||  ||  | |  || | || ||||||   ||| ||| |||
    TGC-GGGAGCTGGCCCTGGGCAGCAACAAGC-TGGGTGATGTGGGCATGGCGGAGCTGTGCCCAGGGCTGCT
      810       820       830       840       850       860       870

        740       750       760       770             780       790
    CCTCCCCTCCCATCCCTAGGATTCGAGCCTTGTGCATC-----GT-TGAC-TCA--GCCAGGGTCTTATCTC
    || ||||  |  |||  ||| |  |  || ||||| |||      || || | |||  |||| || ||   |
    CCACCCCAGC--TCC--AGGCTCAGGACCCTGTGGATCTGGGAGTGTGGCATCACTGCCAAGGGCT--GCGG
     880       890       900       910       920       930       940

      800       810       820       830       840       850
    GG-TTACACCTGTAATCTCA-GCCCTTGG-GAGCC--CAG-AG--CAGGATTGCTG--AATGGTCTGGAGCA
    || |     | ||   |||| | ||  || |||||   || ||  |||  | || |  || |  |||| | |
    GGATCTGTGCCGTGTCCTCAGGGCCAAGGAGAGCCTGAAGGAGCTCAGCCTGGCCGGCAACGAGCTGGGGGA
        950       960       970       980       990      1000      1010

  860       870       X
    GGTCGTCTCGTTCCAGTCGAC
     |  |    || || | |  |
    TGAGG----GTGCCCGACTGCTGTGTGAGAC
          1020      1030      1040


7. US-08-162-407-1 (1-879)
   Q05150       Modified human adipsin gene with leader sequence.

 ID   Q05150 standard; DNA; 835 BP.
 AC   Q05150;
```

```
DT   05-NOV-1990  (first entry)
DE   Modified human adipsin gene with leader sequence.
KW   Adipsin; complement D; obesity; lupus erythrymatosis;
KW   rheumatoid arthritis; ds.
FH   Key             Location/Qualifiers
FT   CDS             11..769
FT   /*tag=  a
FT   misc_RNA        1..84
FT   /*tag=  b
FT   /label=leader sequence
PN   WO9006365-A.
PD   14-JUN-1990.
PF   21-NOV-1989; 005374.
PR   30-NOV-1988; US-277963.
PA   (BETH-) Beth Israel Hospita.
PA   (DANA-) Dana-Farber Cancer Inst.
PA   (META-) Metabolic Biosystems Inc.
PI   Flier JS, Spiegelman BM, Rosen BM, White RT;
DR   WPI; 90-209777/27.
DR   P-PSDB; R05773.
PT   Human protein having adipsin and complement D activity -
PT   used in diagnosis and control of metabolically caused obesity
PT   and for treating infection.
PS   Claim 2; Fig 3; 43pp; English.
CC   Abs raised to the gene product may be used in diagnosis of
CC   metabolically caused obesity. The protein may also be used to
CC   treat and prevent obesity and bacterial, viral, parasitic and
CC   neoplastic cell infection.
SQ   Sequence   835 BP;   128 A;   295 C;   294 G;   118 T;

Initial Score    =    129  Optimized Score  =    270  Significance =  7.47
Residue Identity =    50%  Matches         =    322  Mismatches   =   247
Gaps             =     73  Conservative Substitutions             =     0

  270       280       290       300       310       320       330       340
    AAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATG
              ||| || |  | | || |||| | |  ||   |    | ||  | |||      ||| || |
              GGATCC-CACCAT-GCACAGCTCCGTGTACTTC----GCAG-TTCTGG-TCCTCCTAGGA-G
              X        10        20        30             40         50

         350       360       370       380       390       400
    CAAACGCTTCTGGAGGACGTCAACACCGAGAT-ACAT-TTTGTCACCTCATGTACCTTCCAGC-CCCT--AC
    |   |||  |||     || |  | ||| | |   ||  | | |  |  |   ||     || | ||   |
    CGGCCGC--CTGCGCCGCG-CGGCCCCGTGGTCGGATCCTGGGCGGCAGAGAGGCCGAGGCGCACGCTCGGC
         60        70        80        90       100       110       120

   410       420       430       440       450       460       470
    CAGA-ATGTCTGCGATTCGTCCAGACCAACATCTCCCACCT-CCTGAAGGACACCTG--CACACAGC--TGC
    |  | |||   ||| |  || |||   |||  | | |||||    | |||   ||||    |  |||  ||
    CTTACATG---GCG-TCGGTGCAGCTGAACGGCGCGCACCTGTGCGCAGGCGTCCTGGTGGCGGAGCGGTGG
         130       140       150       160       170       180       190

         480       490       500       510       520       530
    TTGCT----CTGAAGCCCTGTATCG-GGA--AGGCCTGCCAGAATTTCTCTCGGT-GCCTGGAGGTGCA---
     ||||      |  || |||  | | |||   ||||  |  |||  |  |  | |  |||||  | |||
    GTGCTGAGCGCGGCGCACTGCCTGGAGGACGCGGCCGACGGGAAGGT-GCAGGTTCTCCTGGGCGCGCACTC
         200       210       220       230       240       250       260

        540       550       560       570       580       590       600
    ---GTGCCAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCC---CATAGCCCTA-GAAGCCACGGAGC
```



```
      ||  |||||||||  ||||||  |  |||   | | |   |||   ||  |||| |    | ||   |||
    CCTGTCGCAGCCGGAGCCCTCCAAGCGCCTG-TACGACGTGCTCCGCGCAGTGCCCCACCCGGACAGCCAGC
         270       280       290       300       310       320       330

          610       620       630       640       650       660       670
    TCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACT-GGTGCTGCTGGCA
     || | | | ||| ||  |  | ||  |||| || | ||||  |  |      ||||| ||  ||||||
    -CCGACACCATCGACCACG--ACCTCCTGCTGCTACAGCTGTCGGAGAAGGCCACACTGGGCCCTGCTGTGC
          340       350       360       370       380       390       400

          680       690       700       710       720       730
    GCCGCCTGGGGCCTTCGCTG-GC--AAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGGG--GTGCCCCTCC
    ||| |||   ||| | || | ||    |  || |   ||    |||| ||   | ||| |  ||| ||  |
    GCCCCCT---GCCCTGGCAGCGCGTGGACCGCGACGTGGCACCGGGAACT---CTCTGCGACGTGGCCGGCT
             410       420       430       440           450       460

   740       750       760       770       780       790       800
    CCTCCCATCCCTAGGATTCGAGCCTTGTGCATCGTTGACT-CAGC-CAGGGTCTTATCTCGGTTACACC-TG
        |     | |  |  || |||      |       | || |||| |  | ||||  |   | |  ||| |
    GGGGCATAGTCAACCACGCGGGCCGCCGCCCGGACAGCCTGCAGCACGTGCTCTTGCCAGTGCTGGACCGCG
    470       480       490       500       510       520       530

    810       820       830       840       850       860       870
    TAATCT-CAGCCCTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGAGCAGGTCGT-CTCG--TTCCAGT
      | || || ||    || || |  | || ||  ||   ||    |  | || |  | ||  | || | ||
    CCACCTGCAACC----GGCGCACGCACCACGACGGCGCCATCACCGAGCGCTTGATGTGCGCGGAGAGCAAT
  540       550       560       570       580       590       600

     X
    CGAC
    || |
    CGCCGGGACAGCTG
    610       620


8. US-08-162-407-1 (1-879)
   Q03566       Human adipsin/D cDNA hg 31-40.

 ID   Q03566 standard; cDNA; 847 BP.
 AC   Q03566;
 DT   30-JUL-1990  (first entry)
 DE   Human adipsin/D cDNA hg 31-40.
 KW   CAT; hybrid protein; Human adipsin/D;ss.
 OS   Homo sapiens.
 PN   WO9001540-A.
 PD   22-FEB-1990.
 PF   09-AUG-1989; U03417.
 PR   11-AUG-1988; US-231224.
 PA   (CALB-) Calif Biotechn Inc.
 PI   Hilliker S, White R;
 DR   WPI; 90-083499/11.
 DR   P-PSDB; R05421.
 PT   Heterologous protein expression on prokaryotic host -
 PT   using 3' truncated chloramphenicol acetyl transferase gene to
 PT   stably express hybrid protein.
 PS   Example; Fig 10; 67pp; English.
 CC   When inserted into an expression vector, pTrpCAT 72,
 CC   the construct gave 10-15% levels of fusion protein upon
 CC   induction in W3110 cells.
 CC   See also Q03557 to Q05366; and Q04767.
```

```
 SQ   Sequence   847 BP;   132 A;   291 C;   299 G;   125 T;

Initial Score    =    129  Optimized Score  =    261  Significance =  7.47
Residue Identity =    49%  Matches          =    311  Mismatches   =   245
Gaps             =     74  Conservative Substitutions               =     0

      280     X 290       300       310       320       330       340
    TTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATGCAAACG
                |||   |  | | || |   |  |  |  | ||  | |||   | ||  |  ||   ||
              AATTC---GGGCGGGGCAGTTCTGGTCCTCCTAGGA--GCGGC--CGCCTGCGCCGC--GCG
              X           10        20        30          40        50

    350       360       370       380       390       400       410
    CTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCTACCAGAATGTCTG
       |  | || ||  | | | | |   ||    | | |     | ||  ||  | || |||    |||   |
    -GCCCCGTGGTCG-GATC-CTGGGCGGCAGAGAGGC-CGAGGCGCACGCTC--GGCCTTAC----ATG---G
          60          70        80        90        100               110

  420       430       440       450       460       470               480
    CGATTCGTCCAGACCAACATCTCCCACCT-CCTGAAGGACACCTG--CACACAGC--TGCTTGCT----CTG
    || |  || |||   |||  | | |||||    | |||   ||||    |  |||  ||  ||||      |
    CG-TCGGTGCAGCTGAACGGCGCGCACCTGTGCGCAGGCGTCCTGGTGGCGGAGCGGTGGGTGCTGAGCGCG
          120       130       140       150       160       170       180

        490           500       510       520       530               540
    AAGCCCTGTATCG-GGA--AGGCCTGCCAGAATTTCTCTCGGT-GCCTGGAGGTGCA------GTGCCAGCC
      || |||  | | |||   ||||  |  |||  |  |  | |  |||||  | |||      ||  |||||
    GCGCACTGCCTGGAGGACGCGGCCGACGGGAAGGT-GCAGGTTCTCCTGGGCGCGCACTCCCTGTCGCAGCC
        190       200       210       220       230       240       250

       550       560       570       580       590       600       610
    GGACTCCTCCACCCTGCTGCCCCCAAGGAGTCC---CATAGCCCTA-GAAGCCACGGAGCTCCCAGAGCCTC
    |||  ||||||  |  |||   | | |   |||   ||  |||| |    | ||   ||| || | | | ||
    GGAGCCCTCCAAGCGCCTG-TACGACGTGCTCCGCGCAGTGCCCCACCCGGACAGCCAGC-CCGACACCATC
       260       270       280       290       300       310       320

         620       630       640       650       660       670       680
    GGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACT-GGTGCTGCTGGCAGCCGCCTGGGGC
    | ||  |  | ||  |||| || | ||||  |  |     ||||||  ||  ||||||   ||| |||   ||
    GACCACG--ACCTCCTGCTGCTACAGCTGTCGGAGAAGGCCACACTGGGCCCTGCTGTGCGCCCCCT---GC
       330         340       350       360       370       380       390

       690           700       710       720       730       740       750
    CTTCGCTG-GC--AAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGGG--GTGCCCCTCCCCTCCCATCCCT
    | | || | ||    |  || |   ||    ||||  ||   | ||| |  |||  ||  |      |     |
    CCTGGCAGCGCGTGGACCGCGACGTGGCACCGGGAACT---CTCTGCGACGTGGCCGGCTGGGGCATAGTCA
          400       410       420           430       440       450       460

           760       770       780       790       800       810
    AGGATTCGAGCCTTGTGCATCGTTGACT-CAGC-CAGGGTCTTATCTCGGTTACACC-TGTAATCT-CAGCC
    |  |  || |||     |       | || |||| |  | ||||  |   | |  |||  |  | || || ||
    ACCACGCGGGCCGCCGCCCGGACAGCCTGCAGCACGTGCTCTTGCCAGTGCTGGACCGCGCCACCTGCAACC
          470       480       490       500       510       520       530

      820       830       840       850       860       870       X
    CTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGAGCAGGTCGT-CTCG--TTCCAGTCGAC
       || || |  | || ||  ||   ||   |  | ||  |  || | ||      || ||| |
    ----GGCGCACGCACCACGACGGCGCCATCACCGAGCGCTTGATGTGCGCGGAGAGCAATCGCCGGGACAGC
            540       550       560       570       580       590 X     600
```



```
    TG


 9. US-08-162-407-1 (1-879)
    Q05149       Human adipsin gene from the clone phg31.

  ID   Q05149 standard; DNA; 1093 BP.
  AC   Q05149;
  DT   05-NOV-1990  (first entry)
  DE   Human adipsin gene from the clone phg31.
  KW   Adipsin; complement D; obesity; lupus erythrymatosis;
  KW   rheumatoid arthritis; ds.
  FH   Key             Location/Qualifiers
  FT   CDS             1..750
  FT   /*tag=  a
  PN   WO9006365-A.
  PD   14-JUN-1990.
  PF   21-NOV-1989; 005374.
  PR   30-NOV-1988; US-277963.
  PA   (BETH-) Beth Israel Hospita.
  PA   (DANA-) Dana-Farber Cancer Inst.
  PA   (META-) Metabolic Biosystems Inc.
  PI   Flier JS, Spiegelman BM, Rosen BM, White RT;
  DR   WPI; 90-209777/27.
  DR   P-PSDB; R05772.
  PT   Human protein having adipsin and complement D activity -
  PT   used in diagnosis and control of metabolically caused obesity
  PT   and for treating infection.
  PS   Disclosure; .
  CC   Abs raised to the gene product may be used in diagnosis of
  CC   metabolically caused obesity. The protein may also be used to
  CC   treat and prevent obesity and bacterial, viral, parasitic and
  CC   neoplastic cell infection.
  SQ   Sequence   1093 BP;   190 A;   343 C;   383 G;   177 T;

Initial Score    =    128  Optimized Score  =    261  Significance =  7.41
Residue Identity =    49%  Matches          =    312  Mismatches   =   244
Gaps             =     74  Conservative Substitutions               =     0

      280     X 290       300       310       320       330       340
    TTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATGCAAACG
                |||   |  | | || |   |  |  |  | ||  | |||   | ||  |  ||   ||
              AATTC---GGGCGGGGCAGTTCTGGTCCTCCTAGGA--GCGGC--CGCCTGCGCCGC--GCCG
              X           10        20        30          40        50

    350       360       370       380       390       400       410
    CTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCTACCAGAATGTCTG
       |  | || ||  | | | | |   ||    | | |     | ||  |  | ||||||    |||   |
    C--CCCGTGGTCG-GATC-CTGGGCGGCAGAGAGGC-CGAGGCGCAC--GCGCGGCCCTAC----ATG---G
          60          70        80        90        100               110

  420       430       440       450       460       470               480
    CGATTCGTCCAGACCAACATCTCCCACCT-CCTGAAGGACACCTG--CACACAGC--TGCTTGCT----CTG
    || |  || |||   |||  | | |||||    | |||   ||||    |  |||  ||  ||||      |
    CG-TCGGTGCAGCTGAACGGCGCGCACCTGTGCGCAGGCGTCCTGGTGGCGGAGCAGTGGGTGCTGAGCGCG
          120       130       140       150       160       170       180

        490           500       510       520       530               540
    AAGCCCTGTATCG-GGA--AGGCCTGCCAGAATTTCTCTCGGT-GCCTGGAGGTGCA------GTGCCAGCC
```

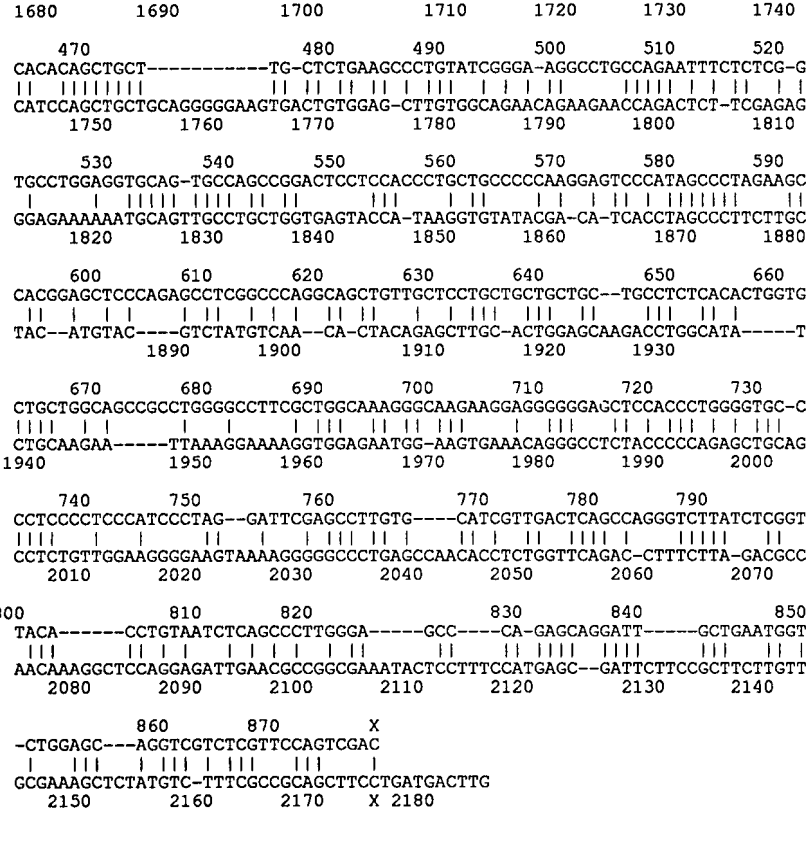
```
GCGCACTGCCTGGAGGACGCGGCCGACGGGAAGGT-GCAGGTTCTCCTGGGCGCGCACTCCCTGTCGCAGCC
    190       200       210       220       230       240       250

  550       560       570       580       590       600       610
GGACTCCTCCACCCTGCTGCCCCCAAGGAGTCC---CATAGCCCTA-GAAGCCACGGAGCTCCCAGAGCCTC
GGAGCCCTCCAAGCGCCTG-TACGACGTGCTCCGCGCAGTGCCCCACCCGGACAGCCAGC-CCGACACCATC
   260       270       280       290       300       310       320

     620       630       640       650       660       670       680
GGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACT-GGTGCTGCTGGCAGCCGCCTGGGGC
GACCACG--ACCTCCTGCTGCTACAGCTGTCGGAGAAGGCCACACTGGGCCCTGCTGTGCGCCCCCT---GC
   330       340       350       360       370       380       390

    690       700       710       720       730       740       750
CTTCGCTG-GC--AAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTG-GGGTGCCCCTCCCCTCCCAT-CCCT
CCTGGCAGCGCGTGGACCGCGACGTGGCACCGGGAACT---CTCTGCGACTTGGCCGGCTGGGGCATAGTCA
      400       410       420       430       440       450       460

       760       770       780       790       800       810
AGGATTCGAGCCTTGTGCATCGTTGACT-CAGC-CAGGGTCTTATCTCGGTTACACC-TGTAATCT-CAGCC
ACCACGCGGGCCGCCGCCCGGACAGCCTGCAGCACGTGCTCTTGCCAGTGCTGGACCGCGCCACCTGCAACC
      470       480       490       500       510       520       530

820       830       840       850       860       870       X
 CTTGGGAGCCCAGAGCAGGATTGCTGAATGGTCTGGAGCAGGTCGT-CTCG--TTCCAGTCGAC
-----GGCGCACGCACCACGACGGCGCCATCACCGAGCGCTTGATGTGCGCGGAGAGCAATCGCCGGGACAGC
          540       550       560       570       580       590 X     600

TG
```

10. US-08-162-407-1 (1-879)
    Q30966 Encodes vitamin K dependent carboxylase.

```
ID   Q30966 standard; DNA; 2452 BP.
AC   Q30966;
DT   25-MAR-1993  (first entry)
DE   Encodes vitamin K dependent carboxylase.
KW   vitamin K dependent carboxylase; factor VII; factor IX; factor X;
KW   protein C; protein S; prothrombin; VKD carboxylase; diagnosis; liver;
KW   cancer; carcinoma; tumour; increased undercarboxylated prothrombin;
KW   ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   CDS             87..2360
FT   /*tag=  a
PN   WO9219636-A.
PD   12-NOV-1992.
PF   08-MAY-1992; U03853.
PR   08-MAY-1991; US-697427.
PR   09-SEP-1991; US-756250.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Stafford DE;
```



```
DR   WPI; 92-398792/48.
DR   P-PSDB; R28954.
PT   DNA coding for vitamin=K-dependent carboxylase - for prodn. of
PT   transformed cells producing carboxylated vitamin=K-dependent
PT   proteins, e.g. clotting factors
PS   Claim 1; Page 36; 55pp; English.
CC   This sequence encodes a vitamin K-dependent carboxylase. The 1.6kb
CC   insert from lambdaZAP-CARB1.6 was used to screen a human cDNA
CC   library. No clones however, coded for the entire sequence of the gamma
CC   glutamyl carboxylase. They did allow comparison of liver carboxylase to
CC   the carboxylase of another tissue. A second cDNA library constructed
CC   using HEL cells (human erythroleukaemia) was screened and a clone
CC   isolated. This contained the entire coding sequence of the human
CC   carboxylase and some upstream and downstream sequences.
SQ   Sequence   2452 BP;    541 A;    654 C;    640 G;    617 T;

Initial Score    =    124  Optimized Score  =    375  Significance =  7.14
Residue Identity =    48%  Matches         =    461  Mismatches   =   363
Gaps             =    135  Conservative Substitutions              =     0

            X          10        20                30        40        50
           GTCGA---CTGGAACGAGACGACCTGCTCT---GT---CACAGGCATGAG-GGGTCCCCGGC
TGGGCTGTATGGCTATTCCTGGGACATGATG--GTGCACTCCCGTTCCCACCAGCACGTGAAGATCACCTAC
     1270 X    1280      1290      1300      1310      1320      1330

       60        70        80        90       100       110       120
AGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCT-CCCTGTTGCTGC--TGTTGCTGCTGCTGAG
CGTGATGGCCGCACTGGCG--AACTGGGCTACCTTAACCCTGGGGTATTTACACAGAGTCGGCGATGGAAGG
     1340      1350      1360      1370      1380      1390      1400

           130       140       150       160           170       180
TCCTTG----CCTGCGGGGGACACCTGACTGTTACTTCAGCCAC---AGTCCCA--TCTCCTCCAACT---T
ATCATGCAGACATGCTGAAGCAATATGCCACTTGCCTGAGCCGCCTGCTTCCCAAGTATAATGTCACTGAGC
     1410      1420      1430      1440      1450      1460      1470

       190       200       210       220       230       240       250
CAAAG-TGAAGTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTC
CCCAGATCTACTTT-GATATTTGGGTCTCCATC-AATGACCGCTT--CCAG-CAGAG-GATTTTTGACCCTC
   1480      1490      1500      1510      1520      1530

      260       270       280       290       300       310       320
AGGACGAGAAGCACTGCAAGGC-CTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACT
GTGTGGACA--TCGTGCAGGCCGCTTGGTCACCCT-TTCAGCGCACATC-CTGGGTGCAACCACTCTTGA-T
1540      1550      1560      1570      1580      1590      1600

     330       340       350       360       370       380       390
GTGGCAGGGT-CTAAGATGCAAACGCTTCTGGAGGACGTCAACA-CCGAGATA--CATTTTGTCACCTCATG
GGACCTGTCTCCCTGGAGGGCCAAGTTACAGGAAATCAAGAGCAGCCTAGACAACCACACTG--AGGTGGTC
1610      1620      1630      1640      1650      1660      1670

      400       410       420       430       440       450       460
TACCTTCCAGCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTG
TTCATTGCAG-ATTTCCCTG---GACTGCACTTGG---AGAATTTTGTGAGTGAAGACCTGGGCAACACTAG
```

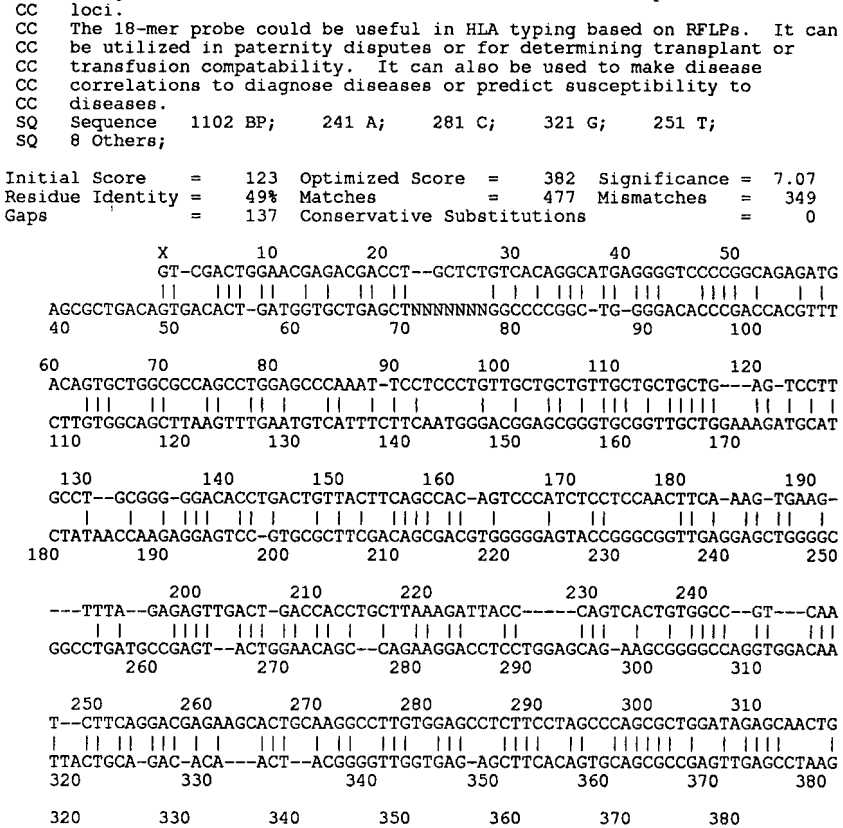


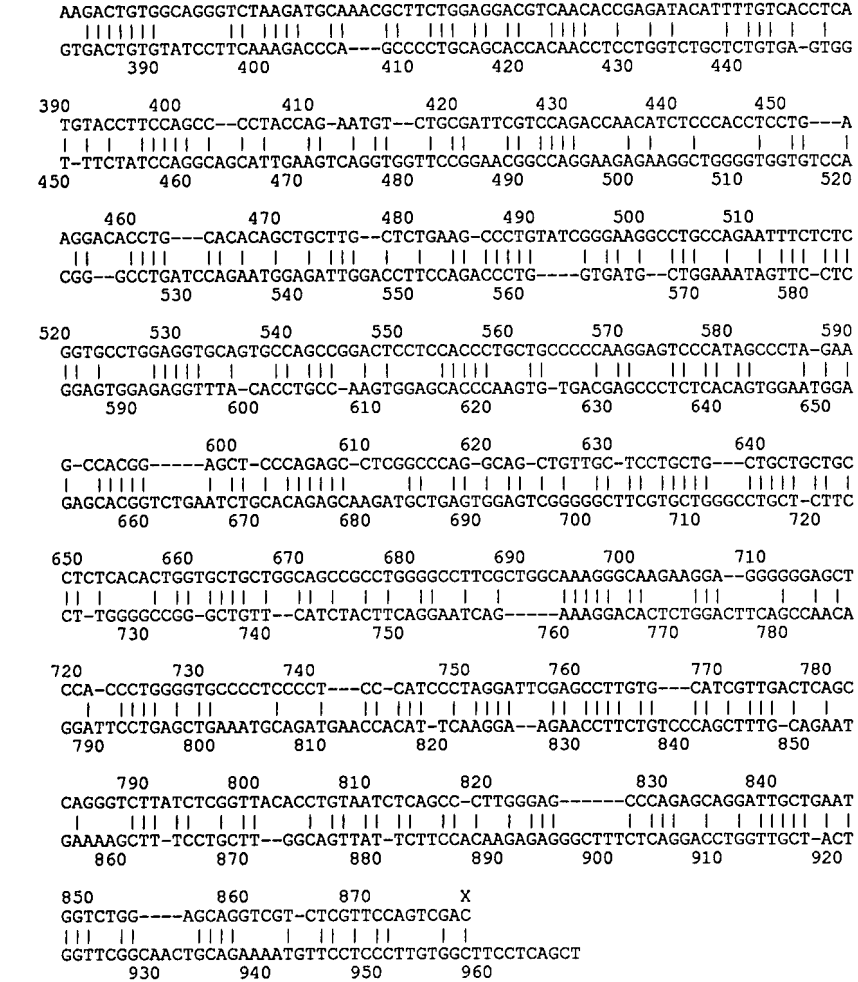
11. US-08-162-407-1 (1-879)
Q25061 Clone 2918.4.

ID Q25061 standard; DNA; 1102 BP.
AC Q25061;
DT 15-NOV-1992 (first entry)
DE Clone 2918.4.
KW RFLP; probe p29G8; beta-lymphoblastoid cell line; LG2; HLA-DRbeta;
KW transplant; transfusion; paternity; Raji cell line; 2918.8; ss.
OS Synthetic.
PN US5110920-A.
PD 05-MAY-1992.
PF 05-DEC-1984; 678255.
PR 22-JAN-1982; US-341902.
PR 07-JAN-1983; US-456373.

PR 05-DEC-1984; US-678255.
PR 30-AUG-1988; US-238619.
PA (CETU) CETUS CORP.
PI Erlich HA;
DR WPI; 92-175244/21.
PT New DNA probes specific to single class II HLA locus - useful in
PT HLA typing e.g. to evaluate paternity and transplant or
PT transfusion compatibility and to diagnose disease susceptibility
PS Disclosure; Page 14; 21pp; English.
CC The sequence given is clone 2918.4 which was derived from a
CC beta-lymphoblastoid cell line LG2 derived cDNA library using an
CC 18-mer probe based on the published amino acid sequence for the
CC conserved amino acids 20-25 of the HLA-DRbeta cDNA clone from the Raji
CC cell line. This probe was hybridized to the cDNA library above
CC and two sequences were isolated, the sequence given and another clone
CC designated 2918.8. These two clones were found to represent different
CC loci.
CC The 18-mer probe could be useful in HLA typing based on RFLPs. It can
CC be utilized in paternity disputes or for determining transplant or
CC transfusion compatability. It can also be used to make disease
CC correlations to diagnose diseases or predict susceptibility to
CC diseases.
SQ Sequence 1102 BP; 241 A; 281 C; 321 G; 251 T;
SQ 8 Others;

Initial Score = 123 Optimized Score = 382 Significance = 7.07
Residue Identity = 49% Matches = 477 Mismatches = 349
Gaps = 137 Conservative Substitutions = 0



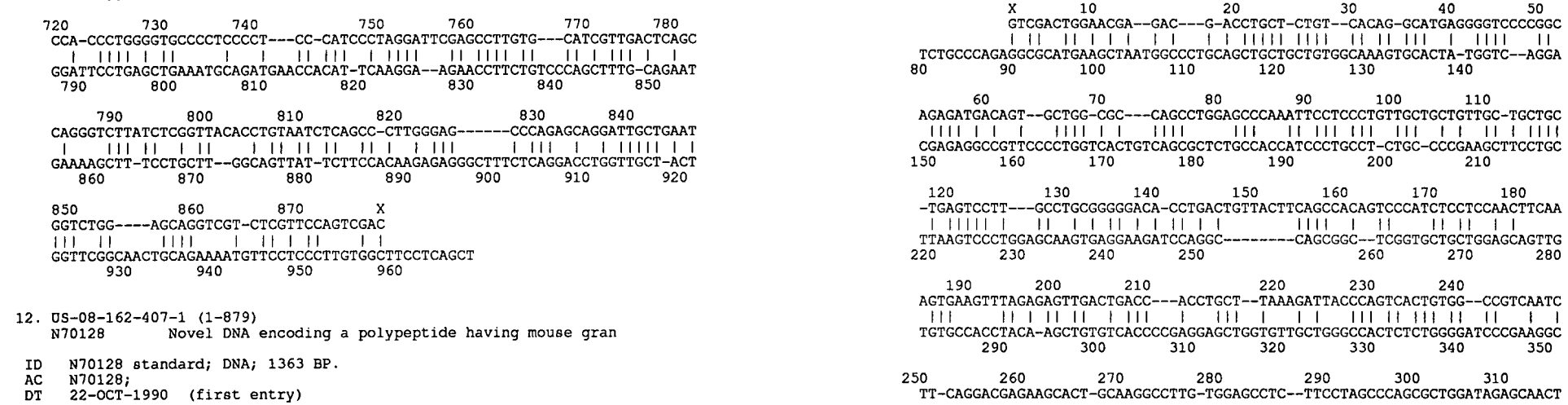


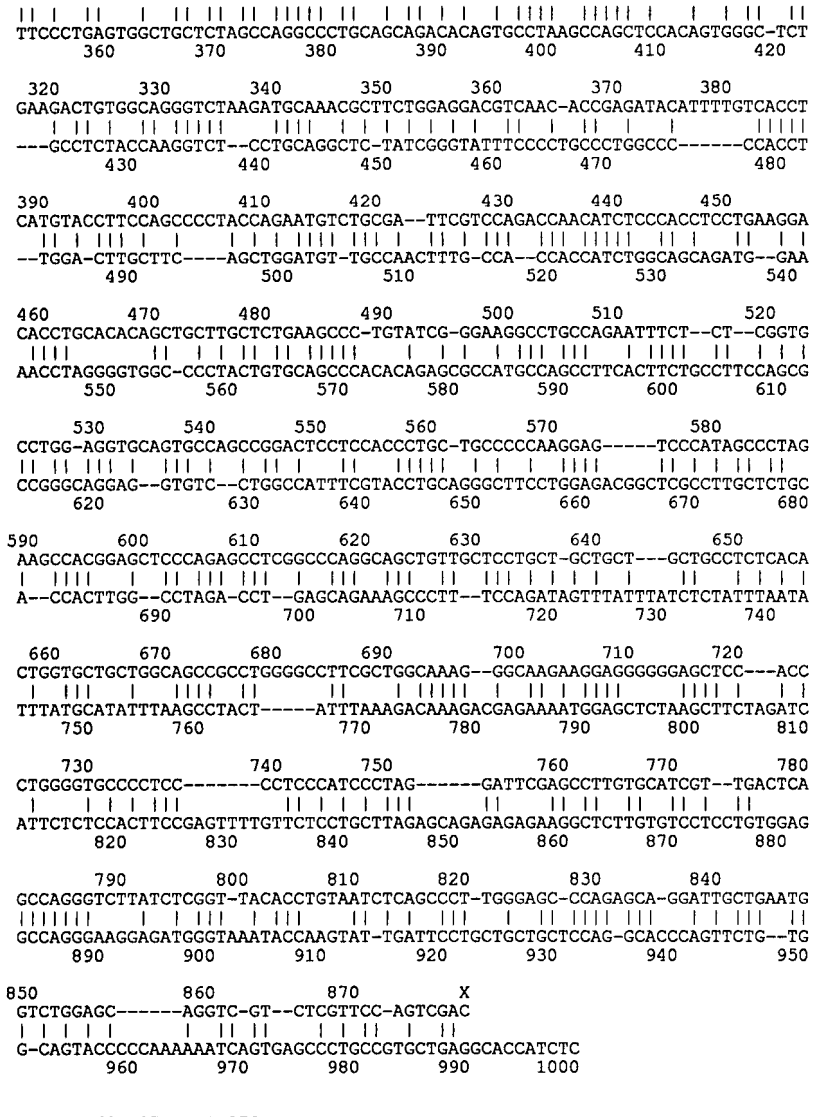
12. US-08-162-407-1 (1-879)
N70128 Novel DNA encoding a polypeptide having mouse gran

ID N70128 standard; DNA; 1363 BP.
AC N70128;
DT 22-OCT-1990 (first entry)

DE Novel DNA encoding a polypeptide having mouse granulocyte
DE colony-stimulating factor (b-CSF) activity is new
KW Mouse granulocyte colony stimulating factor; lymphokine; interleukin.
OS Mouse.
FH Key Location/Qualifiers
FT CDS 68..157
FT /*tag= a
FT /product=Leader peptide
FT mat_peptide 158..694
FT /*tag= b
PN J62269693-A.
PD 24-NOV-1987.
PF 19-MAY-1986; 112506.
PR 19-MAY-1986; JP-112506.
PA (CHUS) Chugai Pharmaceutical Kk.
DR WPI; 88-004545/01.
DR P-PSDB; P70114.
PT New deoxyribonucleic acid -
PT is prepd. by forming mRNA from mammal cells producing
PT polypeptide(s) with mouse granulocyte colony stimulating factor
PT activity
PS Disclosure; Fig 1(A) Page 491; 12pp; Japanese.
CC The CDS for the mature peptide (see FT) is claimed (claims 5 and 6). It
CC was prepd. as follows. mRNA is prepd. from mammal cells capable of
CC producing polypeptides having G-CSF activity and double stranded cDNA is
CC produced from the mRNA by conventional methods. Polypeptides having mouse
CC G-CSF activity are obtd. as 14-75S fractions by the sucrose
CC density-gradient centrifugation method.
SQ Sequence 1363 BP; 279 A; 403 C; 368 G; 313 T;

Initial Score = 123 Optimized Score = 387 Significance = 7.07
Residue Identity = 50% Matches = 488 Mismatches = 325
Gaps = 154 Conservative Substitutions = 0



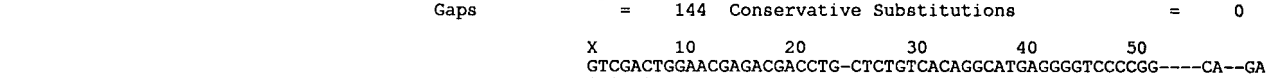


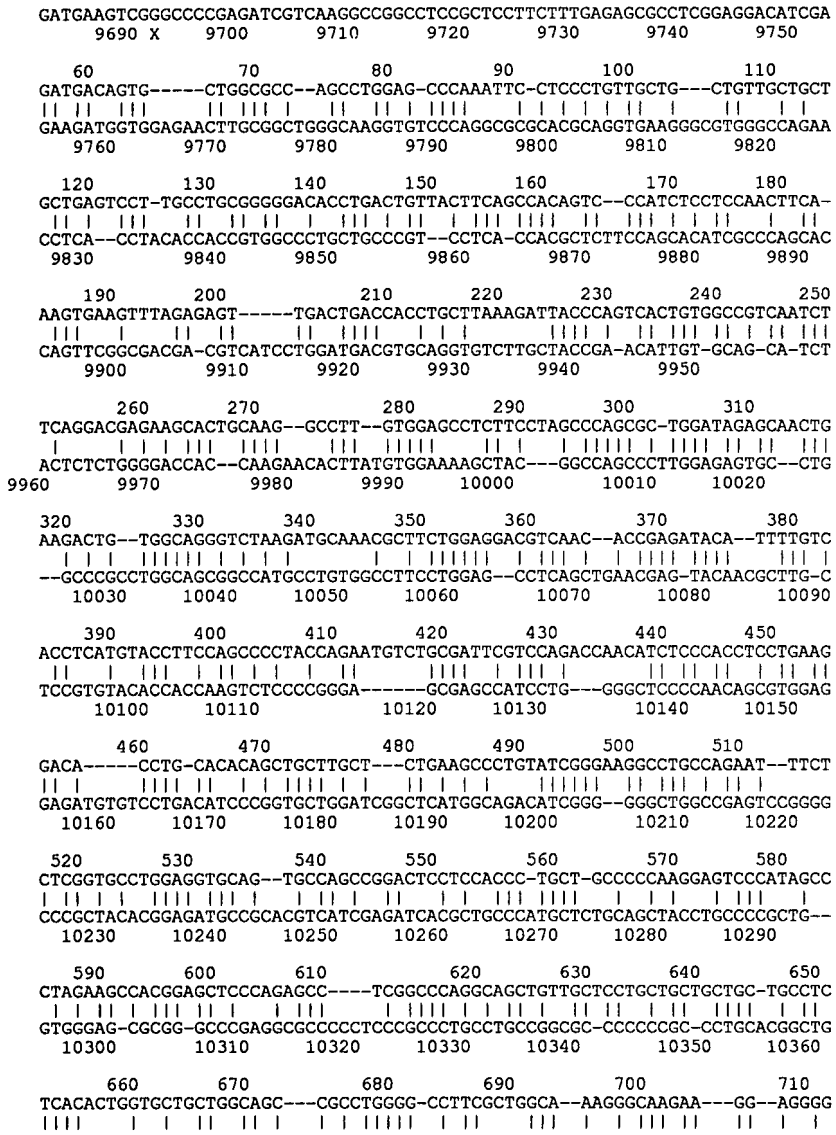
13. US-08-162-407-1 (1-879)

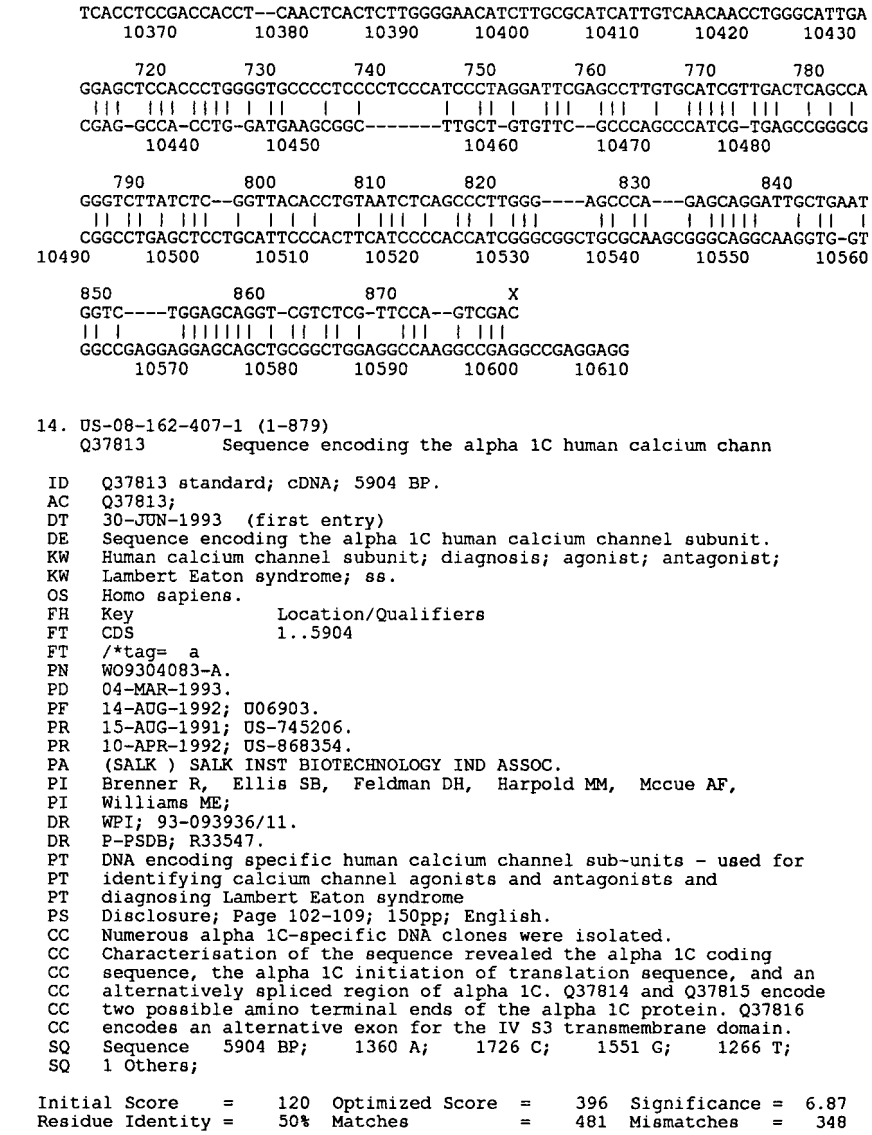
Q10613 Rianodin receptor gene.

```
ID   Q10613 standard; DNA; 15672 BP.
AC   Q10613;
DT   26-APR-1991  (first entry)
DE   Rianodin receptor gene.
KW   Rianodin receptor gene; calcium release modulator; tranquiliser;
KW   antagonist; ss.
OS   Oryctolagus cuniculus.
FH   Key             Location/Qualifiers
FT   CDS             593..15553
FT   /*tag= a
FT   /product= rianodin receptor
FT   GC_signal       351..356
FT   /*tag= b
FT   /note= "feature unlabelled in specification"
FT   GC_signal       407..452
FT   /*tag= c
FT   /note= "feature unlabelled in specification"
FT   CAAT_signal     392..396
FT   /*tag= d
FT   /note= "feature unlabelled in specification"
FT   misc_feature    503..510
FT   /*tag= e
FT   /note= "feature unlabelled in specification"
FT   misc_feature    514..521
FT   /*tag= f
FT   /note= "feature unlabelled in specification"
FT   polyA_signal    15388..15393
FT   /*tag= g
PN   J03011098-A.
PD   18-JAN-1991.
PF   07-JUN-1989; 144569.
PR   07-JUN-1989; JP-144569.
PA   (MITU) Mitsubishi Kasei Corp.
DR   WPI; 91-062003/09.
DR   P-PSDB; R10834.
PT   New rianodin receptor, genes encoding it and its prepn. - useful as
PT   calcium release modulator for tranquillisers and for assaying
PT   calcium antagonists.
PS   Disclosure; Fig 1; 18pp; Japanese.
CC   RNA contg. poly(A) was prepared from rabbit skeletal muscle
CC   endoplasmic reticulum.  From the obtd. poly(A) mRNA, a cDNA bank
CC   corresp. to it was prepared by random primer method, oligo (dT)
CC   primer method, and primer extension method.  A cDNA was obtained by
CC   screening with a DNA probe (see Q10614-15).  By introducing the obtd.
CC   cDNA into an expression vector, vector pRRS7 was formed.
CC   The product is said to be involved in calcium release from
CC   sarcoplasmic reticulum which triggers constriction of skeletal muscle.
CC   Therefore, the receptor is useful as tranquilliser and assay series for
CC   screening of calcium antagonist.
SQ   Sequence   15672 BP;     3139 A;     4909 C;     4955 G;     2669 T;
```

```
Initial Score    =     122  Optimized Score  =    402  Significance =  7.00
Residue Identity =     51%  Matches          =    496  Mismatches   =   326
Gaps             =     144  Conservative Substitutions               =     0
```







14. US-08-162-407-1 (1-879)
Q37813 Sequence encoding the alpha 1C human calcium chann

ID Q37813 standard; cDNA; 5904 BP.
AC Q37813;
DT 30-JUN-1993 (first entry)
DE Sequence encoding the alpha 1C human calcium channel subunit.
KW Human calcium channel subunit; diagnosis; agonist; antagonist;
KW Lambert Eaton syndrome; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..5904
FT /*tag= a
PN WO9304083-A.
PD 04-MAR-1993.
PF 14-AUG-1992; U06903.
PR 15-AUG-1991; US-745206.
PR 10-APR-1992; US-868354.
PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,
PI Williams ME;
DR WPI; 93-093936/11.
DR P-PSDB; R33547.
PT DNA encoding specific human calcium channel sub-units - used for
PT identifying calcium channel agonists and antagonists and
PT diagnosing Lambert Eaton syndrome
PS Disclosure; Page 102-109; 150pp; English.
CC Numerous alpha 1C-specific DNA clones were isolated.
CC Characterisation of the sequence revealed the alpha 1C coding
CC sequence, the alpha 1C initiation of translation sequence, and an
CC alternatively spliced region of alpha 1C. Q37814 and Q37815 encode
CC two possible amino terminal ends of the alpha 1C protein. Q37816
CC encodes an alternative exon for the IV S3 transmembrane domain.
SQ Sequence 5904 BP; 1360 A; 1726 C; 1551 G; 1266 T;
SQ 1 Others;

Initial Score = 120 Optimized Score = 396 Significance = 6.87
Residue Identity = 50% Matches = 481 Mismatches = 348

```
Gaps            =    123  Conservative Substitutions            =     0

            X       10        20        30        40        50
            GTCGACTGGAACGAGACGAC-CTGC-TCTGTCACA--GGCATGAGGGGTCCCCGGCA--GAG
            | | ||   | | |  | || | ||  |||  | |   |||  |    | |   | ||  |
  CGGGGAGGTTGACCACCCCATC-ATCCAACCCCGCAGCTGGGAGAACAGCAAG-TTTGACTTTGACAATGTT
3310      3320      3330      3340      3350      3360      3370

    60          70        80        90       100       110       120
  ATGACAG---TGCTGGCGC-CAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTG--CTGCTGAGT
   || |||   || |||| | |  |   |   |||  || |    || ||    | |||||  | |||   |
  CTGGCAGCCATGATGGCCCTCTTCACCGTCTCCACCTT-CGAAGGGTGGCCAGAGCTGCTGTACCGCTCCAT
3380      3390      3400      3410      3420      3430      3440

        130       140       150       160       170       180       190
  C--CTTGCCTGCGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAGT
  |  ||  |   |||  ||||   | |   | || || | || ||          ||||| ||||    | |  |
  CGACTCCCACACGGAAGACAAGGGCCCCAT-CTACAACTACCGTGTGGAGATCTCCATCTTCTTCATCA--T
3450      3460      3470      3480      3490      3500      3510

      200       210       220       230       240               250
  TTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGC--CGTCA--ATC----TTCAGGA
   || |   |   | | | ||| ||| | | |   |   || |||  |||||   ||    |||||||
  CTACATCAT--CATCATCGCCTTCTTCATGATGAACATCTTCGTG-GGCTTCGTCATCGTCACCTTTCAGGA
3520        3530      3540      3550      3560      3570      3580

        260       270           280       290       300       310
  ------CGAGAAGCACTGCAAGGCCT-TG---TGGAGCCTCTTCCTAGC-CCAGCGC-TGGATAGAGCAACT
        ||| || | | |||| || ||   |||| |     || |||  ||| || |||| | |    ||
  GCAGGGGGAGCAGGAGTACAAGAACTGTGAGCTGGA-CAAGAACC-AGCGACAGTGCGTGGA-ATACGCCCT
3590      3600      3610      3620      3630      3640      3650

  320       330       340       350       360       370       380
  GAAGACTGTGGCAGGGTCTAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTC
   ||| |   |||     |  || | ||     | |   || ||  || ||||   | | | | ||  || |
  CAAGGC-CCGGCCCCTGCGGAGGTACA-----TCCCCAAGAAC--CAGCACCAGTACAAAGTGTGGTACGTG
  3660      3670      3680           3690      3700      3710      3720

390       400       410           420       430       440          450
  ATGTACCTTCCAGCCCCTACCAGAA--TG-TCTGCG-ATTCGTCCAGACCAACATCTCCCACCT--CC-TGA
   |  ||  |||| |  || | || |  || | | ||   || ||| |  ||||| |  |  |||  || ||
  GTCAAC--TCCACCTACTTCGAGTACCTGATGTTCGTCCTCATCCTGCTCAACACCATCTGCCTGGCCATGC
         3730      3740      3750      3760      3770      3780      3790

     460       470       480       490       500       510       520
  AGGACACCTGCACACAGCTGCTTGCTCTGAA--GCCCTG--TATCGGGAAGGCCTGCCAGAATTTCTCTCGG
  || ||  | || || |||||| || ||  ||  ||| ||   |||  ||  | |||       ||| || |
  AGCACTACGGC-CAGAGCTGCCTGTTCAAAATCGCCATGAACATCCTCAA--CATGC----TCTTCACTGGC
        3800      3810      3820      3830      3840          3850

       530       540           550       560       570       580
  TGCCTGGAGGTGCAG-TGCCAGCCGGA--CTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAG----CCCT
    | |    ||| || ||  | || ||   ||| |    ||| |    |||||||  | |  |||    ||||
  CTCTTCACCGTGGAGATG--ATCCTGAAGCTCAT--TGCCTTC-AAACCCAAGGGTTACTTTAGTGATCCCT
   3860      3870      3880      3890      3900      3910      3920

   590       600       610       620       630       640       650
  AGAA-GCCACGGAGCTCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACAC
   ||| |     || || ||  |   |||     ||||||   |  |  ||  | | |  | ||  |  ||| ||
  GGAATGTTTTTGA-CTTCCTCA---TCGTAATTGGCAGC--ATAAT--TGACG-TCATTCT--CAGTCAGAC
```

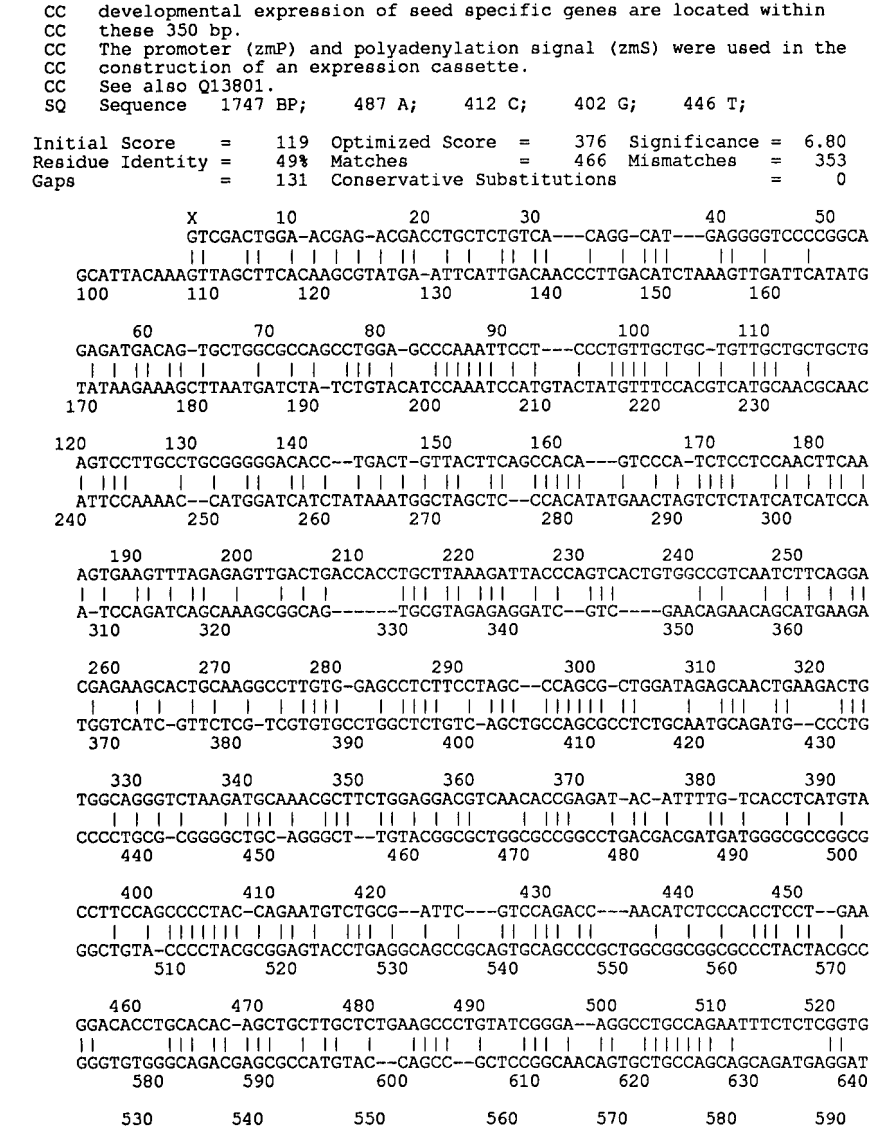


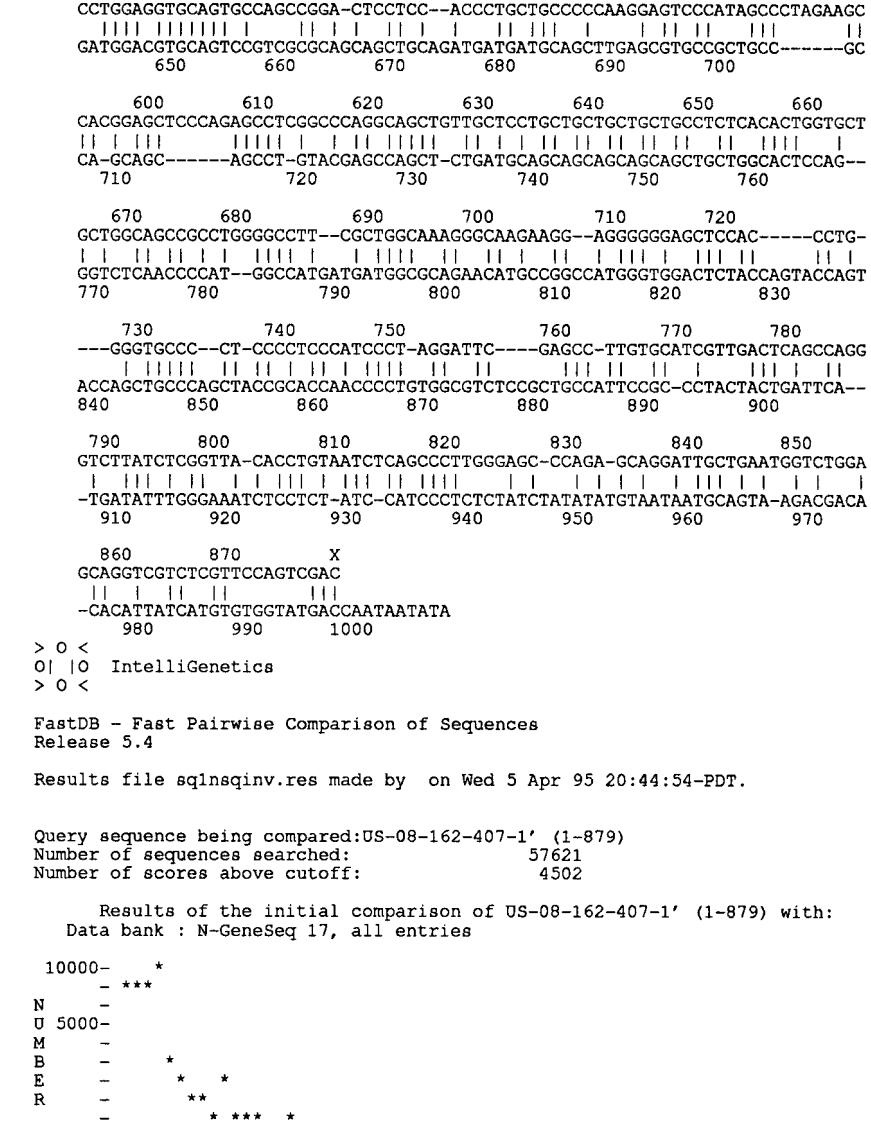
```
      3930      3940      3950      3960      3970      3980
  660       670       680       690       700       710       720
  TGGTGCTGCTG-GCAGCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGAGCTC---CACCCT
  |  | | ||||  ||  | |   |  |   | ||   || |   | |||        | | |||   |||| |
  TAATCCAGCTGAACATACCCAATGCTCTCCCTCTATGAACGCAGAGGAAAACTCCCGCATCTCCATCACCTT
    3990      4000      4010      4020      4030      4040      4050

   730       740       750           760           770       780
  GGGGTGCCCCTCCCCTCCCATCC--CTAG-GATTC----GAGCCTT-----GTGCATCGTTGACTCAGCCAG
      |||  | ||    ||| |  || | ||  |    ||||| |     | |||||   ||| | ||
  CTTCCGCCTGTTCCGGGTCATGCGTCTGGTGAAGCTGCTGAGCCGTGGGGAGGGCATC-CGGACGCTGCTGT
  4060      4070      4080      4090      4100      4110      4120

   790       800       810       820           830       840       850
  GGTCTTATCTCGGTTACACCTGTAATCTCAGCCCT-TG-GGAGC-CCAGAGCAGGATTGCTGAATGGTC-TG
  || | |   ||   |  | | |  | | ||||| || ||  | || || |  || |||||  |  || |
  GGACCTTCATCAAGT--CCTTCCAGGCCCTGCCCTATGTGGCCCTCCTGATCGTGA-TGCTG-TTCTTCATC
  4130      4140      4150      4160      4170      4180      4190

     860       870       X
  GAGCAGGTCGTCTCGTTCCAGTCGAC
   |   |||  ||  | | |||  |
  TACGCGGTGATCGGGATGCAGGTGTTTGGGAAAATT
    4200      4210      4220      4230
```

15. US-08-162-407-1 (1-879)
    Q13798 Zea mays 15 kD seed storage protein gene, pZein15K

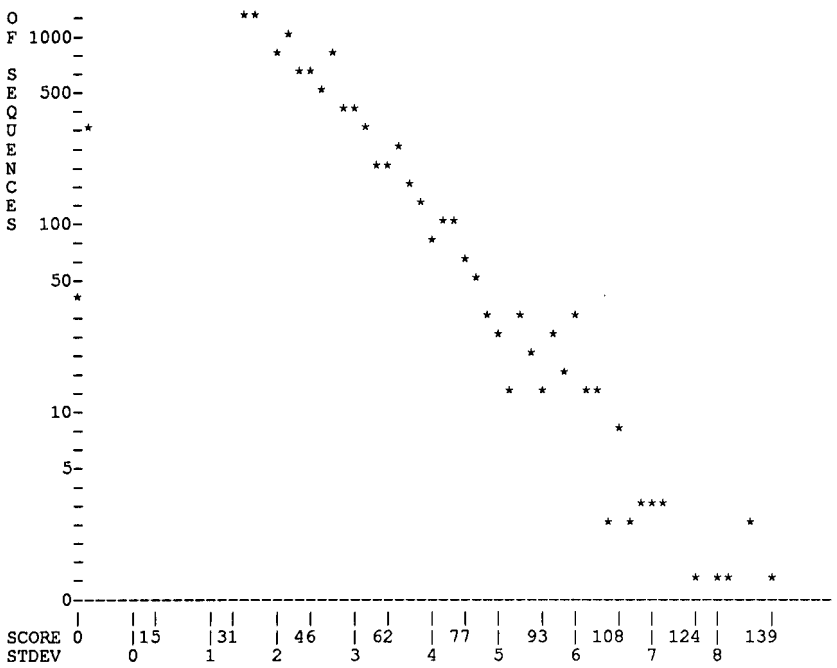
```
ID   Q13798 standard; DNA; 1747 BP.
AC   Q13798;
DT   10-DEC-1991  (first entry)
DE   Zea mays 15 kD seed storage protein gene, pZein15K.
KW   Promoter; beta-zein; polyadenylation signal; zmP; zmS; ss.
OS   Zea mays.
FH   Key             Location/Qualifiers
FT   CDS             360..899
FT   /*tag=  a
FT   /label= seed_storage_protein
FT   /note= "15 kD"
PN   WO9113993-A.
PD   19-SEP-1991.
PF   14-FEB-1991; U00887.
PR   05-MAR-1990; US-489175.
PA   (UPJO ) UPJOHN CO.
PI   Slightom JL, Chee PP;
DR   WPI; 91-295647/40.
DR   P-PSDB; R14108.
PT   Seed specific expression cassettes - having specific gene
PT   regulatory elements for prodn. of proteins in plant seed storage
PT   bodies
PS   Disclosure; Page 28-30; 58pp; English.
CC   The nucleotide sequence of the beta-zein gene reported by Pedersen
CC   only extends ca. 233 bp 5' of the translation initiation.
CC   To obtain more of the beta-zein promoter, a second gene, pZein15K,
CC   was isolated.  The sequence contains an additional 110 bp of 5'
CC   flanking DNA, giving a total promoter region of ca. 350 bp to the
CC   translation initiation codon.  Analysis of many zein type genes
CC   has shown that sequences responsible for controlling the
```

```
CC   developmental expression of seed specific genes are located within
CC   these 350 bp.
CC   The promoter (zmP) and polyadenylation signal (zmS) were used in the
CC   construction of an expression cassette.
CC   See also Q13801.
SQ   Sequence   1747 BP;    487 A;    412 C;    402 G;    446 T;

Initial Score    =    119  Optimized Score  =    376  Significance =  6.80
Residue Identity =    49%  Matches          =    466  Mismatches   =   353
Gaps             =    131  Conservative Substitutions              =     0
```





```
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq1nsqinv.res made by  on Wed 5 Apr 95 20:44:54-PDT.


Query sequence being compared:US-08-162-407-1' (1-879)
Number of sequences searched:                57621
Number of scores above cutoff:                4502

     Results of the initial comparison of US-08-162-407-1' (1-879) with:
   Data bank : N-GeneSeq 17, all entries

 10000-    *
       - ***
N      -
U 5000-
M      -
B      -     *
E      -      *   *
R      -       **
       -         * ***  *
```



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	18	14	14.33

Times:	CPU	Total Elapsed
	00:03:49.01	00:03:49.00

Number of residues:	24347505
Number of sequences searched:	57621
Number of scores above cutoff:	4502

Cut-off raised to 11.
Cut-off raised to 17.
Cut-off raised to 25.
Cut-off raised to 31.
Cut-off raised to 36.
Cut-off raised to 40.
Cut-off raised to 44.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 8 standard deviations above mean ****					
1. Q29269	Human calcium channel 27980/1	6232	139	406	8.44	0
2. Q51731	Plasmid pCisEBON for subcloni	10596	134	385	8.09	0
3. N91772	Human androgen receptor cDNA.	3569	133	399	8.02	0
	**** 7 standard deviations above mean ****					
4. Q43423	Sequence encoding the 10E3B (	739	129	275	7.75	0
5. Q30999	Notch clone hN3k full length	3234	128	272	7.68	0
6. N90369	DNA sequence of alpha-2 plasm	1521	123	355	7.33	0
	**** 6 standard deviations above mean ****					
7. Q04035	Stem cell leukaemia (SCL) gen	4091	117	391	6.91	0
8. Q05330	Stem cell leukaemia (SCL) gen	4199	117	391	6.91	0
9. Q37817	Sequence encoding the alpha 1	7362	116	405	6.84	0
10. N60309	Equine IFN-omega-1 from pAH61	2109	115	340	6.77	0
11. Q03541	Chinese hamster IMPDH.	1620	114	387	6.70	0
12. Q45955	mRXR-beta.	2285	113	386	6.63	0
13. Q43031	Collagen-like polymer DCP2-C2	432	112	204	6.56	0
14. Q43032	Collagen-like polymer DCP3-C2	756	112	315	6.56	0
15. Q37818	Sequence encoding the alpha 1	7175	112	404	6.56	0
16. Q45335	Human TLE-3 gene.	2357	110	389	6.42	0
17. Q62177	Human TLE-3 gene.	2357	110	389	6.42	0
18. Q65395	Osteogenic protein hOP2 gene.	6418	108	370	6.28	0
19. Q13608	ACV synthetase gene, pcbAB.	11601	108	392	6.28	0
20. Q48231	Vector containing ACVS gene.	13058	108	392	6.28	0
21. Q24474	Precursor ET-3 cDNA sequence.	2233	107	391	6.21	0
22. Q11255	Endothelin-3 precursor.	2299	107	396	6.21	0
23. Q11943	Nucleotide sequence of HIV-1	9192	107	373	6.21	0
24. N92768	HIV-2 variant HIV-D194 clone.	9473	107	387	6.21	0
25. N93063	Sequence encoding hybrid prot	1383	106	291	6.14	0
26. Q06029	Sequence encoding recombinant	1774	105	375	6.07	0
	**** 5 standard deviations above mean ****					
27. Q20924	C10-E13 DNA fragment encoding	559	103	243	5.93	0
28. Q23091	Antigen ac-6b gene.	619	103	237	5.93	0
29. N50390	Bovine interferon alpha C.	840	103	306	5.93	0
30. N92097	Combined open reading frames	5300	103	375	5.93	0
31. N90327	Hepatitis C virus composite p	5360	103	375	5.93	0
32. N92103	Combined open reading frames	6905	103	375	5.93	0
33. N92106	Combined open reading frames	7310	103	375	5.93	0
34. N90336	Composite hepatitis C virus (	7310	103	375	5.93	0
35. Q10566	Hepatitis C virus strain 1 DN	9185	103	375	5.93	0
36. Q05956	Sense strand of the compiled	9185	103	375	5.93	0

37.	Q21744	Compiled HCV cDNA.	9400	103	376	5.93	0
38.	Q05955	Hepatitis C virus open readin	8316	102	375	5.86	0
39.	Q12864	Human Cytotoxic Cell Protease	849	101	235	5.79	0
40.	Q05617	Exon 1 of porcine leukaemia i	1140	101	223	5.79	0
41.	Q63870	Feline T cell protein CD4 cDN	1422	101	409	5.79	0
42.	N60077	DNA sequence of the env regio	2607	101	382	5.79	0
43.	N80220	Sequence of Mycobacterium tub	2640	101	317	5.79	0
44.	Q14753	HIV-1 BA-L clone.	3807	101	375	5.79	0
45.	Q43662	Acetobacter cdg3 operon.	4131	101	293	5.79	0

1. US-08-162-407-1' (1-879)
   Q29269 Human calcium channel 27980/11.

```
ID   Q29269 standard; DNA; 6232 BP.
AC   Q29269;
DT   03-MAR-1993  (first entry)
DE   Human calcium channel 27980/11.
KW   Plasmid pR14-5.3.3.1; Ca-flux assay; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   polyA_signal    6215..6220
FT   /*tag= a
FT   repeat_unit     1..252
FT   /*tag= b
FT   /standard_name= Alu_repeat
FT   /note= "possible cloning artefact"
FT   CDS             253..6048
FT   /*tag= c
FT   /note= "amino acids 358 to C-terminus
FT   i.e. Domains II to IV"
FT   misc_difference 3746
FT   /*tag= d
FT   /note= "undefined"
PN   EP-507170-A.
PD   07-OCT-1992.
PF   23-MAR-1992; 104970.
PR   04-APR-1991; DE-110785.
PA   (FARB ) BAYER AG.
PI   Franz J, Rae P, Unterbeck A, Weingaertner B;
DR   WPI; 92-333446/41.
DR   P-PSDB; R27649.
PT   Cloned human neuronal calcium channel sub-types - useful in
PT   calcium flux assays to screen for neurone-specific calcium
PT   channel ligands
PS   Claim 2; Page 63-77; 101pp; German.
CC   Human neuroblastoma cell line, hippocampus, frontal and temporal
CC   cortex and visual cortex cDNA banks were screened with a probe
CC   containing carp skeletal muscle Ca-channel cDNA. The cDNA clone
CC   pR14-5.3.3.1 overlaps with clone p1247-14.1.1.1 (see Q29263). The
CC   following differences are observed between the two sequences
CC   (nucleotide and position in pR14-5.3.3.1 given in brackets):
CC   1.Cytosine at position 520 (T: 3507); no change in deduced amino
CC   acid sequence. 2.Cytosine at position 775 (G; 3768); no change in
CC   deduced AA sequence. 3.Cytosine at position 1617 (T;4611).
CC   4.Adenosine at position 2360 (G; 5353). 5.deletion of 6 nucleotides
CC   at position 708 (CGGAAA; 3695-3700). 6. deletion of an Adenosine
CC   residue at position 1013 which leads to a stop codon at position
CC   1028-1030. 7. at position 3240 there are a further 2199 nucleotides
CC   of the 3'UTR which are absent from pR14-5.3.3.1. (The deletion of
```



```
CC   Adenosine at position 1013 is thought to be a cloning artefact).
CC   The sequence can be inserted into a eukaryotic expression vector for
CC   use in transforming suitable host cells. Cell lines producing human
CC   neuronal calcium channel proteins can be used for screening for Ca
CC   channel ligands (agonists or antagonists). See also Q29259-Q29275.
SQ   Sequence   6232 BP;    1250 A;    1914 C;    1827 G;    1240 T;
SQ   1 Others;
```

Initial Score = 139 Optimized Score = 406 Significance = 8.44
Residue Identity = 51% Matches = 490 Mismatches = 334
Gaps = 122 Conservative Substitutions = 0

```
            X       10        20        30        40        50
            GTCGA-CTGGAACGAGACGACCTGCTCCAGACCAT-TCAG-CA-ATCCTGCTCTG-GGCTCC
            || || |||| || |||| | || | || | || || | || | ||| |||| |
    GCCGCTGGTGGTGGAGCTGGGCCGCGACGGCGCGCGGGGGCCCGTGGGAGGCAAAGCCCGACCTGAGGCTGC
      1480 X   1490      1500      1510      1520      1530      1540

     60        70        80        90       100       110       120
    CAAGGGCTGAGATTACAGGTGTAACC-GAGATAAG--ACC-CTGGCTGAGTCAACGATGCACAAG-GCTCGA
      ||| |   ||   |  ||  | ||   |   ||  ||| | |||   | | || | | |||||  | |
    GGAGGCCCCCGAGGGC--GTCGACCCTCCGCGCAGGCACCACCGGCACCG-CGAC-AAGGACAAGACCCCCG
      1550      1560      1570      1580      1590      1600      1610

       130       140       150       160       170       180       190
    ATCCTAGGGA-TGGGA--GGGGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCCCTTTGCCAGCG
       |  |||| ' |||  | | || ||| ||| | || ||||  |       ||     |||||      ||||
    CGGCGGGGGACCAGGACCGAGCAGAGGC-CCCGAAGGCGGAGAGCGGGGGAGCC-CGGTGCCCGGGAGGAGCG
      1620      1630      1640      1650      1660      1670      1680

        200       210       220       230       240       250       260
    AAGGC-CCCAGGCGGCTGCCAGCAGC-ACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCAACAGCTGCCTG
       ||  || | | || |||| |||| |  ||   | | |||  | | ||  || ||||| |  || |||
    GCCGCGGCCCGACCGCAGCCA-CAGCAAGGAGGCCGCGGGGCCCCCGGAGGCGC-GGAGCGAGCGCGGCCGA
      1690      1700      1710      1720      1730      1740      1750

        270       280       290       300       310       320       330
    GGCCGAGGCTCTGGGAGC-TCCG-TGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAG-GGTGGAGGAG
    |||| |||| | | | ||  |||  |||  |  |||  | | |||| || ||  || |  | |  ||||
    GGCCCAGGCCCCGAGGGCGGCCGGCGGCACCACCGGC---GCGGCTCCCCGGAGGAGGCGGCCGAGCGGGAG
      1760      1770      1780      1790      1800      1810      1820

        340       350       360       370       380       390
    TCCGGCTGGCACTGCACCTCCAGGCACCGAGAGAAATTCTGGC-AGGCCTTC-CCGATAC--AGGGC-TTCA
     || |  | ||| || |   | |||||| || |  || |  || ||| | | |||   |  |||||   |
    CCCCGACGCCACCGCGCGCACCGGCACC-AG-G--ATCCGAGCAAGGAGTGCGCCGGCGCCAAGGGCGAGCG
       1830      1840      1850      1860      1870      1880

  400       410       420       430       440       450       460
    GAGCAAGCAGCTGTGTGCAGGTGTCCTTCAGGAG--GTGGGAGATGTTGGTCTGGACGAATCGCAGACATTC
    | ||  || ||   | |  ||   ||    || |    |||||  |  |  | ||  ||   || | |
    GCGCGCGCGGCACCGCGGCGGCCCCCGAGCGGGGCCCCGGGAGGCGGAGAGCGGG--GAGGAGCCGGCGCGG
1890      1900      1910      1920      1930      1940      1950

  470       480       490       500       510       520       530       540
    TGGTAGGGGCTGGAAGGTACATGAGGTGACAAAATGTATCTCGGTGTTGACGTCCTCCAGAAGCGTTTGCAT
     || |  ||  |  || |||  ||| | ||   ||  || ||  | ||       |  ||||| |    ||
    CGGCACCGG--GCCCGGCACA--AGGCG-CAGCCTG-CTCACGAGGCTG------TGGAGAAG-GAGACCA-
1960      1970      1980      1990      2000           2010
```

```
          550         560        570        580        590                600
  CTTAG-ACCCTGCCAC--AGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAGGAAGAG--GCT------CCAC
  |  || |    |||||  ||       | ||   ||        ||  | | ||||| ||  |||       ||||
  CGGAGAAGGAGGCCACGGAGAAGGAGGCTGAGATAGTGGAAGC-CGACAAGGAAAAGGAGCTCCGGAACCAC
 2020       2030       2040       2050       2060       2070       2080

       610        620        630        640        650        660        670
  AAGGCCTTGCAGTGCTTCT-CGTCCTGAAGATTGACGGCCACAGTGAC--TGGGTAATCTTTAAGCAGGTGG
   || ||  | ||     ||  | |||| |||   |  |  || |||||  |||||    |  |  ||  ||
  CAGCCCCGGGAGCCACACTGTGACCTGGAGA-CCAGTGGGACTGTGACTGTGGGTCCCATGCACACA-CTGC
2090       2100       2110       2120       2130       2140       2150

       680        690        700        710        720        730
  TCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGACTGTGGCTGAAGTAACAGTCAG----G
   ||| || || ||| |          |||| |||||  ||   | | | || || ||  ||| ||||
  CCAG-CACCTGTCTCCA---------GAAGGTGGAG--GAACAGCCAGAGGATGCAG--ACAATCAGCGGAA
2160       2170               2180       2190       2200       2210

740        750            760        770        780            790
  TGTCCCCCGCA--GGCA---AGGACTCAGCAGCAGCAACAGCA--GCA------ACAGGGAGGAATTTGGGC
   ||| | ||||  ||||   ||  | ||| | | | ||||  |  | |       ||| || |     ||||
  CGTCACTCGCATGGGCAGTCAGCCCCCAG-ACCCG-AACACTATTGTACATATCCCAGTGATGCTGACGGGC
  2220       2230       2240       2250       2260       2270       2280

 800        810        820            830        840        850
  TCCAGGCTGGCGCCAG-CACTGTC-ATCTCTG-----CCGGGGACCCCTCATGCCTGTGACAG----AGCAG
   |    || | |  || ||| |||  || | |      || |||||    | |||    | ||    || ||
  CC---TCTTGGGGAAGCCACGGTCGTTCCCAGTGGTAACGTGGACCTGGAAAGCCAAGCAGAGGGGAAGAAG
     2290       2300       2310       2320       2330       2340       2350

860        870        X
  GTCGTCTCGTTCCAGTCGAC
  |  |  | |   | |  |||
  GAGG--TGGAAGCGGATGACGTGATGAGGA
     2360       2370  X   2380
```

2. US-08-162-407-1' (1-879)
Q51731 Plasmid pCisEBON for subcloning huHGF variants.

```
ID   Q51731 standard; DNA; 10596 BP.
AC   Q51731;
DT   31-MAY-1994  (first entry)
DE   Plasmid pCisEBON for subcloning huHGF variants.
KW   Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW   proteolysis resistant; liver; malignancy; CMV-driven;
KW   Cytomegalovirus; episomal expression plasmid; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   enhancer        1..611
FT   /*tag= a
FT   /note= "CMV enhancer/promoter"
FT   promoter        758..775
FT   /*tag= b
FT   /label= SP6_promoter
FT   misc_feature    845..849
FT   /*tag= c
FT   /note= "SP6 RNA start"
FT   misc_feature    902..966
FT   /*tag= d
```



```
FT   /function= cloning_linker
FT   polyA_signal    967..1107
FT   /*tag= e
FT   /note= "SV40 poly A"
FT   misc_feature    1108..1531
FT   /*tag= f
FT   /function= SV40_origin
FT   misc_feature    1580..4189
FT   /*tag= g
FT   /label= EBNA-1
FT   misc_feature    4190..6374
FT   /*tag= h
FT   /function= oriP
FT   repeat_region   4295..4887
FT   /*tag= i
FT   /note= "family of repeats"
FT   misc_structure  5866..5978
FT   /*tag= j
FT   /note= "dyad region"
FT   terminator      6375..6457
FT   /*tag= k
FT   /label= HSV_TK_terminator_3'-end
FT   CDS             6975..7975
FT   /*tag= l
FT   /phenotype= neomycin_resistance
FT   /note= "Tn5 neomycin phosphotransferase gene"
FT   promoter        7975..8112
FT   /*tag= m
FT   /label= TK_promoter
FT   misc_feature    8114..8594
FT   /*tag= n
FT   /function= M13_ori
FT   misc_RNA        8595..10414
FT   /*tag= o
FT   /label= delta_2a
PN   WO9323541-A.
PD   25-NOV-1993.
PF   17-MAY-1993; U04648.
PR   18-MAY-1992; US-884811.
PR   18-MAY-1992; US-885971.
PA   (GETH ) GENENTECH INC.
PI   Godowski PJ,  Lokker NA,  Mark MR;
DR   WPI; 93-386573/48.
PT   Hepatocyte growth factor variants - are resistant to proteolytic
PT   cleavage into its two-chain form, used to treat malignancies
PT   associated with HGF receptor
PS   Example 1; Fig 6; 87pp; English.
CC   Plasmid pCisEBON (a pRK5 derivative) is an episomal CMV driven
CC   expression plasmid. HuHGF variants with enhanced receptor binding
CC   activity were produced by site-directed mutagenesis. Stable
CC   populations of preferred HGF variants were obtained by transfecting
CC   human embryonic kidney 293 cells and then these were subcloned in
CC   pCisEBON. See R52940-R52949 for examples of pref. HGF variants.
SQ   Sequence   10596 BP;    2625 A;    2571 C;    3024 G;    2376 T;

Initial Score    =    134  Optimized Score  =    385  Significance =  8.09
Residue Identity =    48%  Matches          =    461  Mismatches   =   379
Gaps             =    108  Conservative Substitutions               =     0

             X       10         20         30         40         50         60
```

```
            GTCGACTGGAACGAG-ACGACCTGCTCCA-GACCATTCAGCAATCCTGCTCTGGGCTCCCAA
              ||  || | ||| | ||    |   || || ||   || |          ||||    |
GGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCA-GGAGGAGGGGCAGGAGGGGC-AGGAG
     2260 X    2270       2280       2290       2300       2310       2320

        70         80          90         100        110        120
GGGCTGAGATTACAGGTGTA-ACCGAG--ATAAGACCCTG-GCTGAGTCA-ACGATGCACAAGGCTC-GAAT
|||| |      |||| | |    |||    | || | | | | | ||   |  |||  |||  | | |
GGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG
     2330       2340       2350       2360       2370       2380       2390

 130        140        150        160        170        180        190
CCTAGGGATGGG-AGG-GGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCC--CTTTGCCAGCGA
|    ||| ||| ||| | ||| | |      || || ||     |          | |      | ||| ||
CAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAG-GA
   2400       2410       2420       2430       2440       2450       2460

       200        210        220        230        240        250
AGGCC----CCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCA---GCAGCAGCAGGAGCAACAGCTG
 || |      |||| ||  |   |||| | |||   ||| ||||| |   ||||  |||||||| |   || |
GGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
  2470       2480       2490       2500       2510       2520       2530

260        270        280        290        300        310        320        330
CCTGGGC-CGAGGCTCTGGGAGCTCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAGGGTGGAGG
   ||||  ||||  |  |||||   | ||     ||||  |  |||  |    |||||| |||| | |||
GAGGGGCAGGAGGGGC-AGGAGC-AGGAGG-----AGGGGCA--GGA-GCAGGAGGGGCAGGAGGG-GCAGG
  2540       2550       2560             2570       2580       2590

       340        350        360        370        380        390        400
AGTCCGGCTGGCACTGCACCTCCAGGCACCGAGAGAAATTCTGGCAGGCCTTCCCGATACAGG-GCTTCAGA
||   ||| ||  | | |    |||| | |  ||| |      | |||        |   |||| ||   ||
AG--GGGCAGGAGCAGGAGGGGCAGG-AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGG
   2600       2610       2620       2630       2640       2650       2660

      410        420        430        440        450        460
GCAAGCAGCTG--TGTGCAGGTGTCCTTCAGGAGGTG--GGAG-ATGTTGGTCTGGACG---AATCGCA-GA
|  || ||| |   | ||||| |     |||||| |  |||| | |  || | ||| |    |   ||| ||
GGCAGGAGCAGGAGGGGCAGGAG-----CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGA
  2670       2680            2690       2700       2710       2720       2730

   470        480             490        500        510        520
CATTCTGGTAGGGGCTGGA-----AGGTACATGAGGTG-ACAAAATGTATCTCGGTGTTGACGTCC---TCC
    | || |||||| |||      |||  || |||| | |  |   | |    ||    || | |     |
GGGGCAGG-AGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGC
      2740       2750       2760       2770       2780       2790       2800

530        540        550        560        570        580        590
AGAAGCGTTTGCATCTTAGAC-CCTGC--CACAGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAGGAAGAGG
|| ||     |||      | |    ||   | | |   ||  ||   |  ||| |  |||| ||| || ||
AGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA-GCAG-GAGGGGC-AGG-AGGGG
    2810       2820       2830       2840       2850       2860       2870

 600        610        620        630        640        650        660
CTCCACAAGGCCTTGCA-GTGCTTCTCG-TCCTGAAGATTGACGGCCACAGTGACTGGGTAATCTTTAAGCA
|   |  |||     ||| | |   |  |  |  || ||  | | |   ||| ||  ||| |        ||||
CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAG-GA-GGGGCA-----GGAGCA
      2880       2890       2900       2910       2920              2930

 670        680        690        700        710        720        730
```

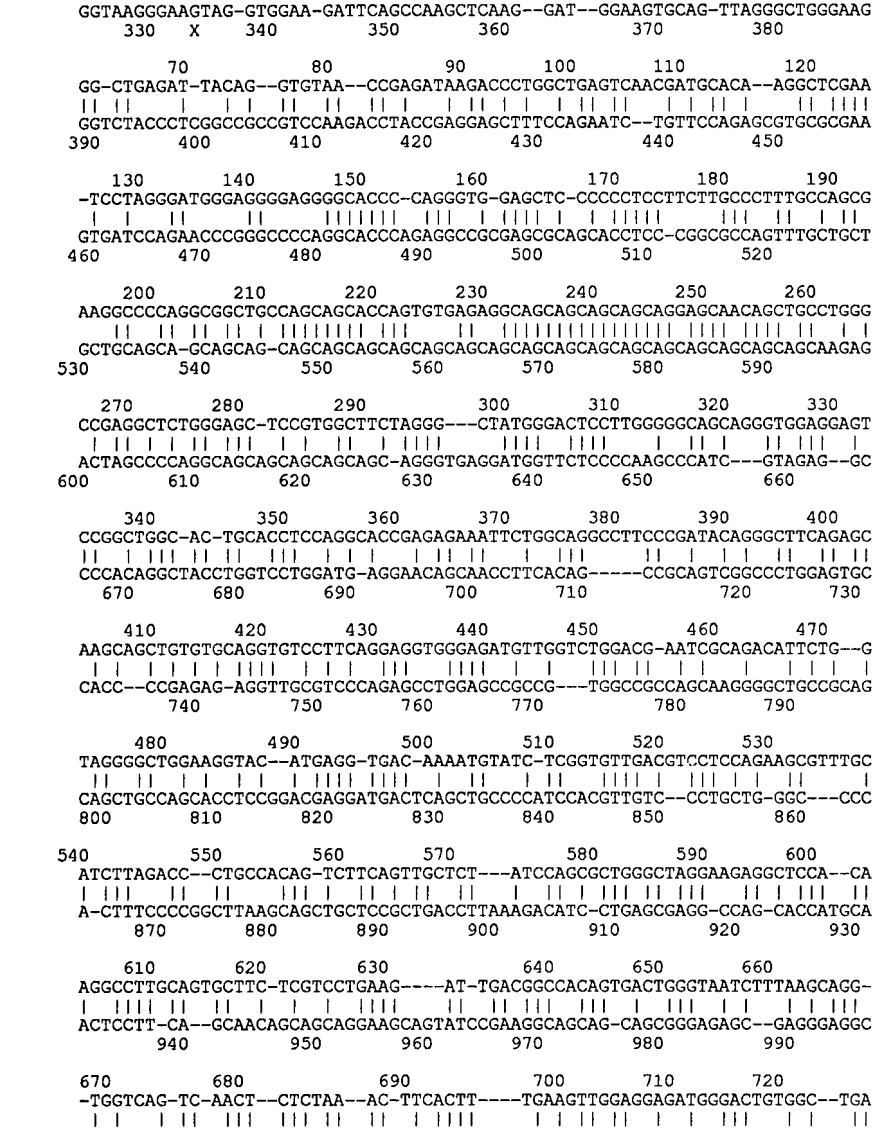


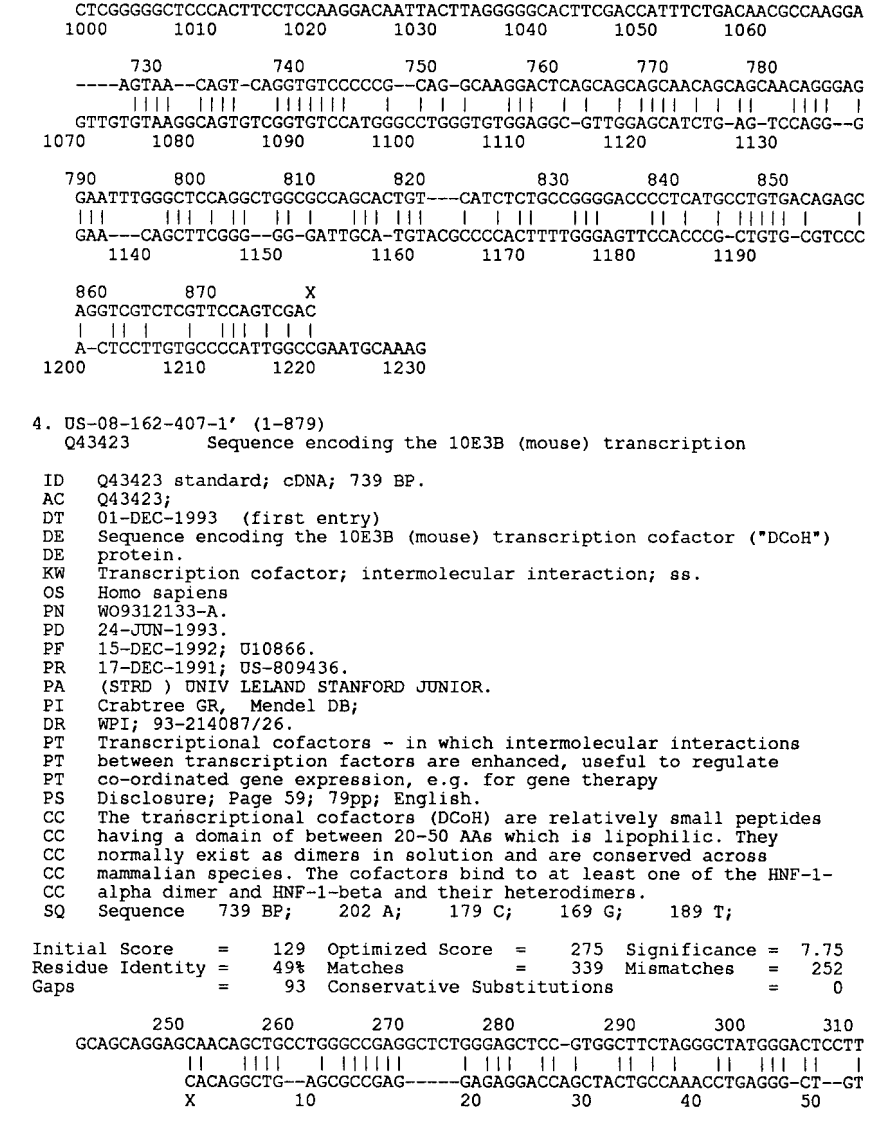
```
   GGTGGT-CAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAG--GAGATGGGACT-GTGGCTGAAGT--AAC
   || |||  || |     || |     ||  | || |  || |  |||  || | | | ||| |  ||  |
   GGAGGTGGAGGCCGGGGTCGAGGAGGCA-GTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGG
    2940      2950       2960       2970       2980       2990       3000

          740        750        760        770        780        790        800
   AGTCA--GGTGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAA--CAGGGAGGAATTTGGGCTC
   ||  |  || | || |||  |   |||||   | |  ||| |  |  | ||  |  | | ||     |||
   AGGTAGTGGAGGCCGCCGGGGTAGAGGAC-GTGAAAGAGCCAGGGGGGGAAGTCGTGAAAGAGCCAGGGGGA
   3010       3020       3030       3040       3050       3060       3070

                  810            820        830        840        850
   CAGG--------CTGGCGCCA-----GCACTGTCATCTCTGCCG-GGGAC-CCCTCATGCCTGTGACAGAGCA
     |||        ||| |  |     || | |   || | |  |   | |  | |||| |  ||  |  | |
   GAGGTCGTGGACGTGGAGAAAAGAGGCCCAGGAGTCCCAGTAGTCAGTCATCATCATCCGGGTCTC-CACCG
  3080       3090       3100       3110       3120       3130       3140

  860        870         X
   GGTCGTCTCGTTCCA-GTCGAC
    |  | | |  |||| || ||
   CGCAGGCCCCCTCCAGGTAGAAGGCCATTTTT
 3150       3160       3170       3180
```

3. US-08-162-407-1' (1-879)
   N91772 Human androgen receptor cDNA.

```
ID   N91772 standard; cDNA; 3569 BP.
AC   N91772;
DT   19-MAR-1990  (first entry)
DE   Human androgen receptor cDNA.
KW   Human androgen receptor; monoclonal antibody; ploYclonal antibody;
KW   cancer; probe.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             363..3122
FT   /*tag=  a
PN   WO8909791-A.
PD   19-OCT-1989.
PF   13-APR-1989; U01548.
PR   14-APR-1988; US-182646.
PA   (UYNC-) University of North Carolina.
PI   French FS, Wilson EM, Joseph DR, Lubahn DB;
DR   WPI; 89-324206/44.
DR   P-PSDB; P93109.
PT   DNA encoding androgen receptor protein - useful for transforming
PT   eukaryotic hosts for protein expression and subsequent antibody prodn.
PS   Disclosure; Fig. 4; 41pp; English.
CC   Complementary DNA sequences derived from the cDNA may be used as probes
CC   detect the presence of androgen receptor (AR mRNA in tumour cells, and to
CC   detect AR gene defects using DNA hybrisidation assays.
SQ   Sequence   3569 BP;   796 A;   1008 C;   975 G;   790 T;
```

```
Initial Score    =    133  Optimized Score  =    399  Significance =  8.02
Residue Identity =    52%  Matches          =    495  Mismatches   =   323
Gaps             =    130  Conservative Substitutions               =     0

              X       10         20         30         40         50         60
             GTCGACTGGAACGAGAC-GACCTGCTCCAGACCATTCAGCAATCCTGCTCTGGGCTCCCAAG
             || |  ||||| ||  | | |  |||| ||   ||   | | | | | |  |||||   |||
```

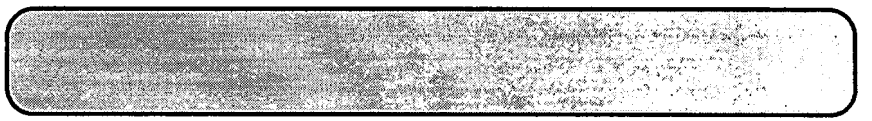


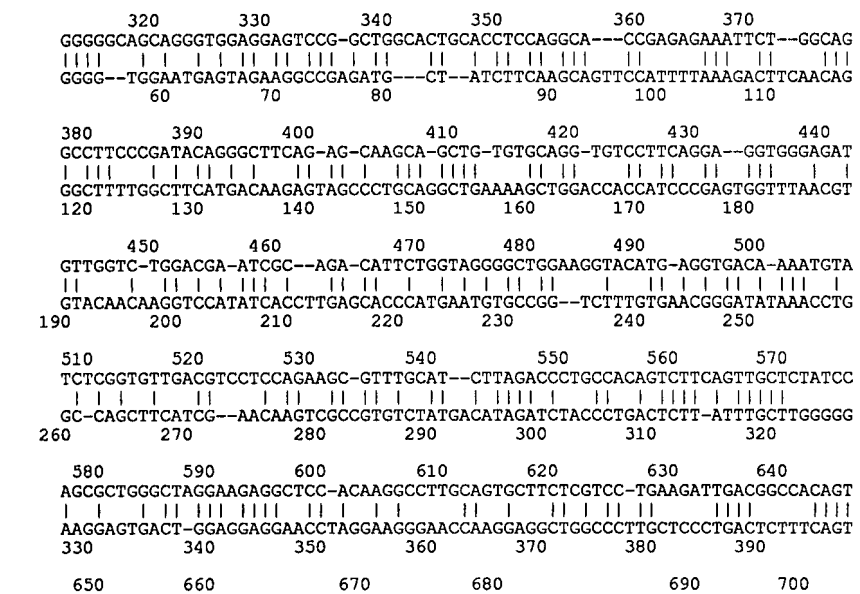


4. US-08-162-407-1' (1-879)
Q43423 Sequence encoding the 10E3B (mouse) transcription

ID Q43423 standard; cDNA; 739 BP.
AC Q43423;
DT 01-DEC-1993 (first entry)
DE Sequence encoding the 10E3B (mouse) transcription cofactor ("DCoH")
DE protein.
KW Transcription cofactor; intermolecular interaction; ss.
OS Homo sapiens
PN WO9312133-A.
PD 24-JUN-1993.
PF 15-DEC-1992; U10866.
PR 17-DEC-1991; US-809436.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Crabtree GR, Mendel DB;
DR WPI; 93-214087/26.
PT Transcriptional cofactors - in which intermolecular interactions
PT between transcription factors are enhanced, useful to regulate
PT co-ordinated gene expression, e.g. for gene therapy
PS Disclosure; Page 59; 79pp; English.
CC The transcriptional cofactors (DCoH) are relatively small peptides
CC having a domain of between 20-50 AAs which is lipophilic. They
CC normally exist as dimers in solution and are conserved across
CC mammalian species. The cofactors bind to at least one of the HNF-1-
CC alpha dimer and HNF-1-beta and their heterodimers.
SQ Sequence 739 BP; 202 A; 179 C; 169 G; 189 T;

Initial Score = 129 Optimized Score = 275 Significance = 7.75
Residue Identity = 49% Matches = 339 Mismatches = 252
Gaps = 93 Conservative Substitutions = 0



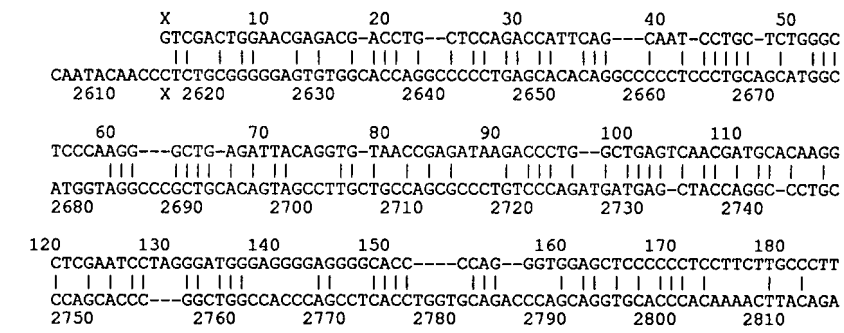


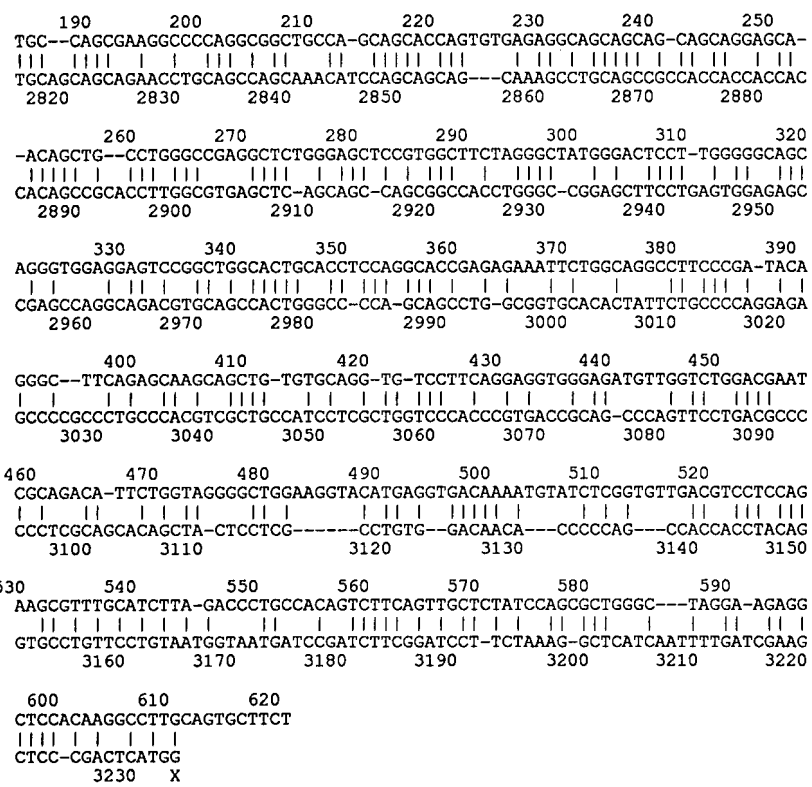
5. US-08-162-407-1' (1-879)
Q30999 Notch clone hN3k full length clone.

ID Q30999 standard; DNA; 3234 BP.
AC Q30999;

DT 01-APR-1993 (first entry)
DE Notch clone hN3k full length clone.
KW Human; Notch; plasmid; cDNA; clone; hN2k; expression library;
KW adhesion domain; Drosophilla; epidermal growth factor; EGF;
KW EGF-like; homologous; repeat; ELR; ELR-11; ELR-12; PCR;
KW polymerase chain reaction; primer; cloning vector; Delta; Serrate;
KW neurogenic; toporythmic; homotypic; heterotypic; differentiation;
KW quantitation; antibody; ss.
OS Homo sapiens.
PN WO9219734-A.
PD 12-NOV-1992.
PF 01-MAY-1992; U03651.
PR 03-MAY-1991; US-695189.
PR 14-NOV-1991; US-791923.
PA (INDV) UNIV INDIANA FOUND.
PA (UYYA) UNIV YALE.
PI Artavanis-Tsakonas S, Blaumueller CM, Fehon RG, Muskavitch MAT;
PI Rebay I, Shepard SB;
DR WPI; 92-398861/48.
PT Human Notch and Delta DNA and protein sequences - used for study
PT and manipulation of differentiation processes
PS Claim 132; Fig 23; 239pp; English.
CC The sequence given represent the full length nucleotide sequence of
CC human Notch contained in plasmid cDNA clone hN3k. A human expression
CC library was constructed and screening assays were carried out on the
CC basis of positive binding to the adhesion domain of Notch, ie. the
CC region that is homologous to Drosophilla epidermal growth factor
CC (EGF)-like (homologous) repeat (ELR)-11 and -12. Alternatively the
CC sequence could be isolated by amplification using polymerase chain
CC reaction (PCR) primers. The isolated gene may be inserted into a
CC cloning vector and expressed. The Notch gene and also the Delta and
CC Serrate neurogenic genes are designated "toporythmic" genes. The
CC proteins they encode are involved in specific homo- or heterotypic
CC interactions crucial to differentiation. The quantitation of mRNA
CC for human Notch and Delta and adhesive molecules, and study of its
CC expression are possible using the DNA and antibodies raised against
CC the Notch and Delta proteins.
SQ Sequence 3234 BP; 608 A; 1130 C; 999 G; 497 T;

Initial Score	=	128	Optimized Score	=	272	Significance	= 7.68
Residue Identity	=	49%	Matches	=	326	Mismatches	= 252
Gaps	=	75	Conservative Substitutions				= 0



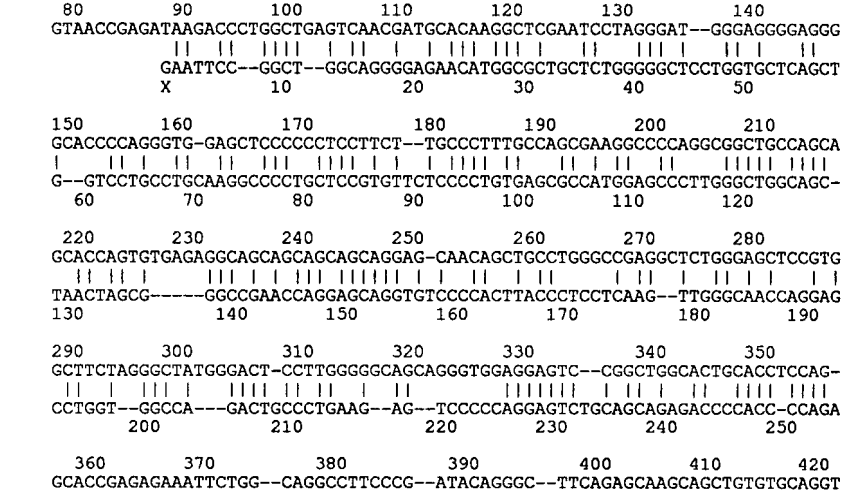


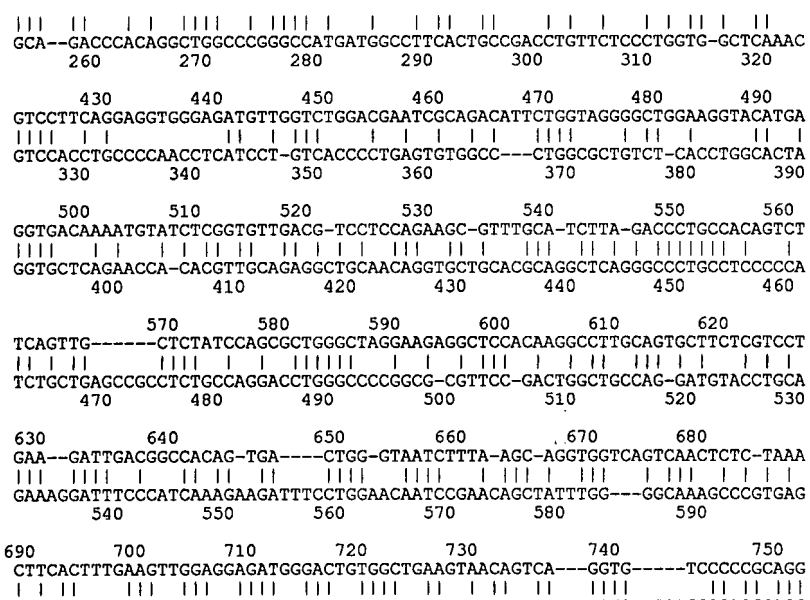
6. US-08-162-407-1' (1-879)
N90369 DNA sequence of alpha-2 plasmin inhibitor deriv.

ID N90369 standard; DNA; 1521 BP.
AC N90369;
DT 18-OCT-1989 (first entry)
DE DNA sequence of alpha-2 plasmin inhibitor deriv.
KW Alpha-2 plasmin inhibitor deriv.; antithrombin-active
KW serine protease inhibitor; thrombin.
OS Escherichia coli
FH Key Location/Qualifiers
FT CDS 142..1497
FT /*tag= a
FT misc_feature 1..6
FT /*tag= b
FT misc_feature 7..24

FT /*tag= c
FT misc_signal 25..141
FT /*tag= d
PN EP-236013-A.
PD 02-AUG-1989.
PF 19-JAN-1989; 100851.
PR 28-JAN-1988; JP-15762.
PA Hoechst Japan Ltd.
PI Toba M, Tone M, Kikuno R, Hashimoto T;
DR WPI; 89-221985/31.
PT Alpha-2 plasmin inhibitor deriv. - having reactive site cleavable with
PT plasmin deleted or replaced to give protein active in inhibiting blood
PT coagulation.
PS Disclosure; ; 16pp; English.
CC The DNA encodes an alpha-2 plasmin inhibitor (API) deriv., with the
CC normal API reactive site cleavable with plasmin deleted or replaced
CC with that of antithrombin active serine protease inhibitor cleavable with
CC thrombin. It is active in inhibiting blood coagulation, and can be used
CC in preventitive treatment or therapy of cardiovascular disturbances or
CC diseases. The DNA sequence is expressed in plasmid p-alpha-AP216.
CC Misc. feature b is the EcoRI site from the EcoRI linker used in the
CC cloning; c is a non-coding region upstream of the 5' end; and d
CC corresponds to the peptide contd. only in the alpha-2-PI (the signal
CC peptide coding portion or anchor portion). When the host is an animal
CC cell, the gene carrying the anchor portion is incorporated into the
CC vector, but when host is prokaryotic, this gene is deleted, and an ATG
CC codon is added, then incorporated. See also P90534 for encoded peptide,
CC and P90537.
SQ Sequence 1521 BP; 296 A; 503 C; 435 G; 287 T;

Initial Score = 123 Optimized Score = 355 Significance = 7.33
Residue Identity = 50% Matches = 426 Mismatches = 318
Gaps = 102 Conservative Substitutions = 0



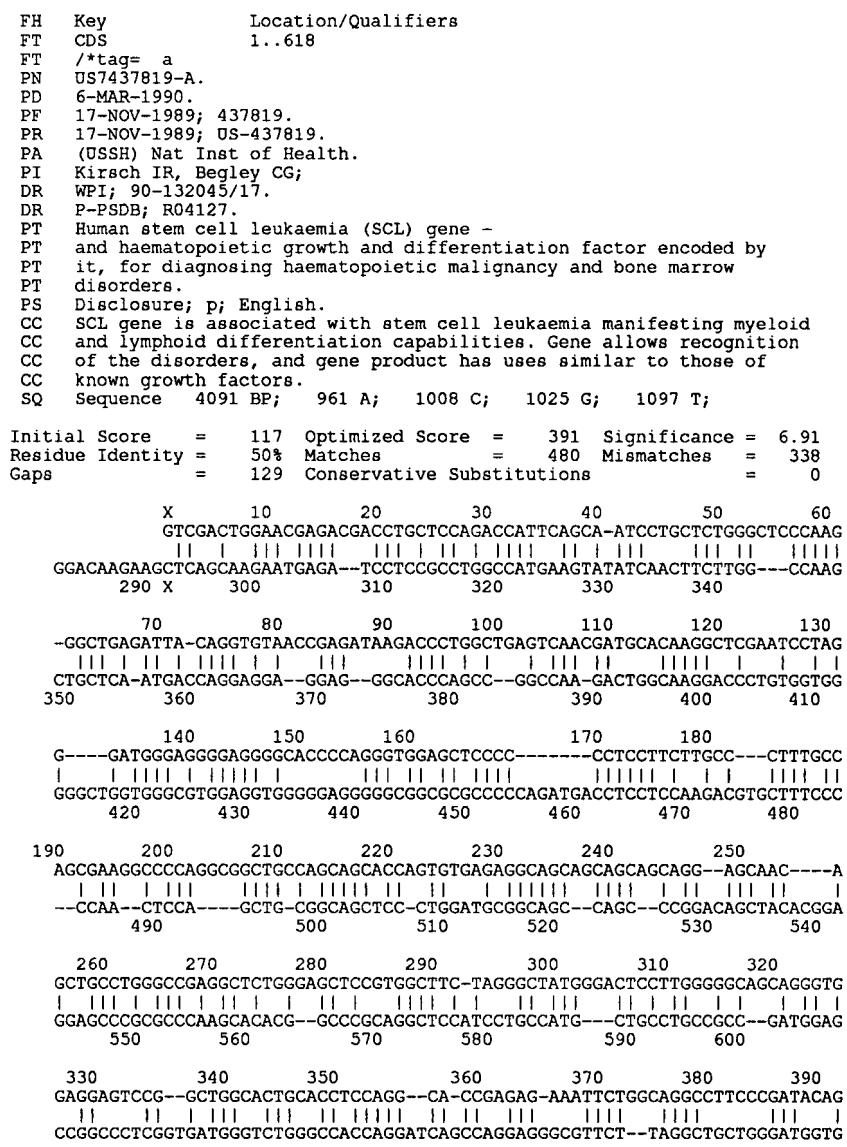


7. US-08-162-407-1' (1-879)
Q04035 Stem cell leukaemia (SCL) gene.

ID Q04035 standard; cDNA; 4091 BP.
AC Q04035;
DT 10-SEP-1990 (first entry)
DE Stem cell leukaemia (SCL) gene.
KW Stem cell leukaemia; leukaemia; SCL; AML; ALL;
KW haematopoietic; aplastic anaemias; ds.
OS Homo sapiens.

FH Key Location/Qualifiers
FT CDS 1..618
FT /*tag= a
PN US7437819-A.
PD 6-MAR-1990.
PF 17-NOV-1989; 437819.
PR 17-NOV-1989; US-437819.
PA (USSH) Nat Inst of Health.
PI Kirsch IR, Begley CG;
DR WPI; 90-132045/17.
DR P-PSDB; R04127.
PT Human stem cell leukaemia (SCL) gene -
PT and haematopoietic growth and differentiation factor encoded by
PT it, for diagnosing haematopoietic malignancy and bone marrow
PT disorders.
PS Disclosure; p; English.
CC SCL gene is associated with stem cell leukaemia manifesting myeloid
CC and lymphoid differentiation capabilities. Gene allows recognition
CC of the disorders, and gene product has uses similar to those of
CC known growth factors.
SQ Sequence 4091 BP; 961 A; 1008 C; 1025 G; 1097 T;

Initial Score = 117 Optimized Score = 391 Significance = 6.91
Residue Identity = 50% Matches = 480 Mismatches = 338
Gaps = 129 Conservative Substitutions = 0



```
610        620        630        640        650        660        670

      400              410        420        430        440        450
GGCTTCAGAGCAAG------CAGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGGTCTGGAC-GAA
GGCTTCAGGGCAGGTGGGGTCAGAATTGGGCGGCTCTCAAGCA--AGGCGGTCGA-CTTCAACTTTCCTGGA
680        690        700        710        720        730        740

460        470        480        490        500        510        520
TCGCAGACATTCTGGTAGGGGCTGGAAGGTACATGAGG-TGAC-AAAATGTATCTCGGTGTTG-ACGTCCTC
TGTCTGA-ACTTTGGGA-AGCCTTTACTGACCCTGGGGCTGGCTTTTCTGTTTCCTGTACCAGTACGAGAT-
 750        760        770        780        790        800        810

 530        540              550        560        570        580
CAGAAGCGT-TTGCA---TCTTAG--AC--CCTGCCACAGTCTTCAGTTGCTCTATCCAGCGCTGGGCT-AG
CAGAAAAATGGAGCAAAGTGGTAGGTACTTTTTGTGA-AGACGGCA--CGGTCT-TCC--CTCTTCCCTCAG
  820        830        840        850        860              870        880

590        600        610        620        630        640        650
--GAAGAGGCTCCACAAGGCCTTGCAGTGCT--TCTCGTC-CTGAAGATTGAC-GGCCACAGTGA-CTGGGT
TCCCAAATCCTTCCCAAG---TAAGAG-GCTGGAGTTGTCACTGCTTTTGGCCTGGAGTTTGGGATCCCTGT
      890        900        910        920        930        940

   660        670        680        690        700        710        720
AATCTTTAAG-CAGGTGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGACTGTGG--C
CTTTCCTAAGACCTGGGGT-TGTCAGCTCTC--ATCTGAGGCATCCAGCAGTCTCTGCCTTGCCTTTAGCCC
950        960        970        980        990       1000       1010

     730        740        750        760        770        780        790
TGAAGTAACAGTCAGGTGTCCCCCG-CAGGCAAGGACTCAGCAGCAGCAACAGCA-GCAACAGGGAGGAATT
CTCCCAAGCTGGCTGGGGTGGCCTGTGTGGGCACTTCTGTCCATATTTATAGGTACCCAATAGCTGCCCATT
1020       1030       1040       1050       1060       1070       1080       1090

         800        810        820        830        840        850
---TGGGCTCCAGGCTGGCGCCAGCACTGTCAT-CTCTGCCGGGGACCCCTCATGCCTGTGACAGAG---CA
TCGTGAGCCCCA-TCTTCACCCAGGCCTATGTTGATCCATCCAGCTTGCCAGATG-CTG---CAGAGTCACA
       1100       1110       1120       1130       1140       1150

 860        870        X
GGTCGTCTCGTTCCAGTCGAC
AG-CCTCGAGGTGCCTTCTTCAGGGCCTGGT
1160       1170       1180
```

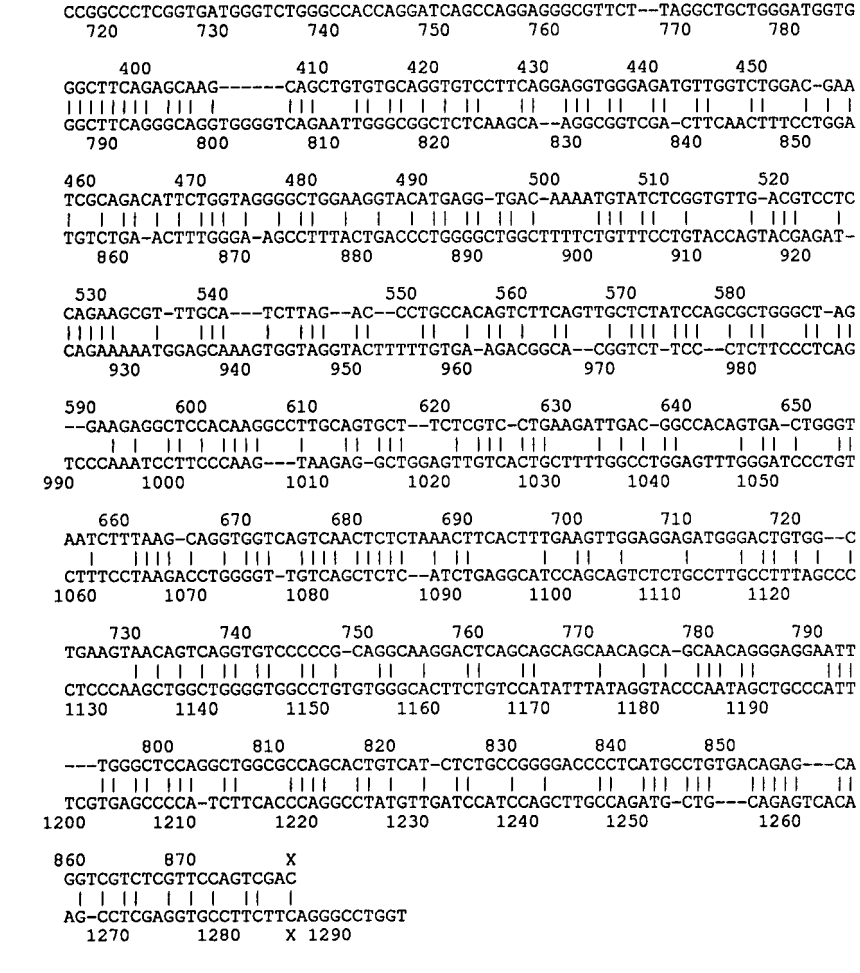
8. US-08-162-407-1' (1-879)
Q05330 Stem cell leukaemia (SCL) gene.

ID Q05330 standard; DNA; 4199 BP.
AC Q05330;
DT 10-SEP-1990 (first entry)
DE Stem cell leukaemia (SCL) gene.
KW Stem cell leukaemia; leukaemia; SCL; AML; ALL;
KW haematopoietic; aplastic anaemias; ds.

OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 85..726
FT /*tag= a
PN US7437819-A.
PD 6-MAR-1990.
PF 17-NOV-1989; 437819.
PR 17-NOV-1989; US-437819.
PA (USSH) Nat Inst of Health.
PI Kirsch IR, Begley CG;
DR WPI; 90-132045/17.
DR P-PSDB; R05857.
PT Human stem cell leukaemia (SCL) gene -
PT and haematopoietic growth and differentiation factor encoded by
PT it, for diagnosing haematopoietic malignancy and bone marrow
PT disorders.
PS Disclosure; p; English.
CC SCL gene is associated with stem cell leukaemia manifesting myeloid
CC and lymphoid differentiation capabilities. Gene allows recognition
CC of the disorders, and gene product has uses similar to those of
CC known growth factors.
SQ Sequence 4199 BP; 966 A; 1052 C; 1036 G; 1137 T; 8 Others;

Initial Score = 117 Optimized Score = 391 Significance = 6.91
Residue Identity = 50% Matches = 480 Mismatches = 338
Gaps = 129 Conservative Substitutions = 0

```
            X        10        20        30        40        50        60
            GTCGACTGGAACGAGACGACCTGCTCCAGACCATTCAGCA-ATCCTGCTCTGGGCTCCCAAG
  GGACAAGAAGCTCAGCAAGAATGAGA--TCCTCCGCCTGGCCATGAAGTATATCAACTTCTTGG---CCAAG
390        400        410        420        430        440        450

          70         80         90        100        110        120        130
-GGCTGAGATTA-CAGGTGTAACCGAGATAAGACCCTGGCTGAGTCAACGATGCACAAGGCTCGAATCCTAG
CTGCTCA-ATGACCAGGAGGA--GGAG--GGCACCCAGCC--GGCCAA-GACTGGCAAGGACCCTGTGGTGG
 460        470        480        490        500        510        520

          140        150        160              170        180
G----GATGGGAGGGGAGGGGCACCCCAGGGTGGAGCTCCCC-------CCTCCTTCTTGCC---CTTTGCC
GGGCTGGTGGGCGTGGAGGTGGGGGAGGGGGCGGCGCGCCCCCAGATGACCTCCTCCAAGACGTGCTTTCCC
        530        540        550        560        570        580        590

190        200        210        220        230        240        250
AGCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGG--AGCAAC----A
--CCAA--CTCCA----GCTG-CGGCAGCTCC-CTGGATGCGGCAGC--CAGC--CCGGACAGCTACACGGA
          600        610        620        630        640        650

   260        270        280        290        300        310        320
GCTGCCTGGGCCGAGGCTCTGGGAGCTCCGTGGCTTC-TAGGGCTATGGGACTCCTTGGGGGCAGCAGGGTG
GGAGCCCGCGCCCAAGCACACG--GCCCGCAGGCTCCATCCTGCCATG---CTGCCTGCCGCC--GATGGAG
        660        670        680        690        700        710

 330        340        350        360        370        380        390
GAGGAGTCCG--GCTGGCACTGCACCTCCAGG--CA-CCGAGAG-AAATTCTGGCAGGCCTTCCCGATACAG
```

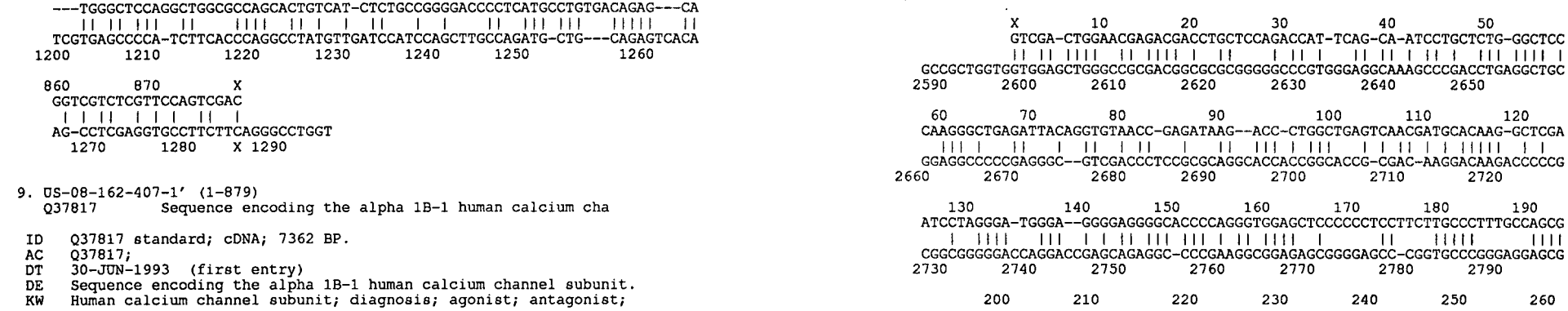


9. US-08-162-407-1' (1-879)
Q37817 Sequence encoding the alpha 1B-1 human calcium cha

ID Q37817 standard; cDNA; 7362 BP.
AC Q37817;
DT 30-JUN-1993 (first entry)
DE Sequence encoding the alpha 1B-1 human calcium channel subunit.
KW Human calcium channel subunit; diagnosis; agonist; antagonist;

KW Lambert Eaton syndrome; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 144..7163
FT /*tag= a
PN WO9304083-A.
PD 04-MAR-1993.
PF 14-AUG-1992; U06903.
PR 15-AUG-1991; US-745206.
PR 10-APR-1992; US-868354.
PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,
PI Williams ME;
DR WPI; 93-093936/11.
DR P-PSDB; R33549.
PT DNA encoding specific human calcium channel sub-units - used for
PT identifying calcium channel agonists and antagonists and
PT diagnosing Lambert Eaton syndrome
PS Disclosure; Page 111-120; 150pp; English.
CC DNA encoding the alpha 1B subunit was isolated by screening a
CC human basal ganglia cDNA library with fragments of the rabbit
CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.
CC A portion of one of the positive clones was used to screen an IMR32
CC cell cDNA library. Clones that hybridized to the basal ganglia
CC DNA prove were used to further screen an IMR32 cell cDNA library
CC to identify overlapping clones that in turn were used to screen a
CC human hippocampus cDNA library. In this way, a sufficient series of
CC clones to span nearly the entire length of the nucleotide sequence
CC encoding the human alpha 1B subunit was obtained. PCR amplification
CC of specific regions of the IMR32 cell alpha 1B mRNA yielded
CC additonal segments of the alpha 1B coding sequence. A full-length
CC alpha 1B DNA clone was constructed by ligating portions of the
CC partial cDNA clones (see Q37817, Q37818). Alpha 1B-1 and alpha
CC 1B-2 are derived by alternative splicing of the alpha 1B subunit
CC transcript.
SQ Sequence 7362 BP; 1446 A; 2276 C; 2214 G; 1426 T;

Initial Score = 116 Optimized Score = 405 Significance = 6.84
Residue Identity = 51% Matches = 491 Mismatches = 331
Gaps = 126 Conservative Substitutions = 0



```
---AAGGCCCCAGGCGGCTGCCAGCAGC-ACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCAACAGCTGCC
GCCGCGGCCGCA--CCGCAGCCA-CAGCAAGGAGGCCGCGGGGCCCCCGGAGGCGC-GGAGCGAGCGCGGCC
2800        2810        2820        2830        2840        2850        2860

      270         280         290         300         310         320         330
TGGGCCGAGGCTCTGGGAGC-TCCG-TGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAG-GGTGGAGG
GAGGCCCAGGCCCCGAGGGCGGCCGGCGGCACCACCGGC---GCGGCTCCCCGGAGGAGGCGGCCGAGCGGG
 2870        2880        2890        2900        2910        2920        2930

       340         350         360         370         380         390
AGTCCGGCTGGCACTGCACCTCCAGGCACCGAGAGAAATTCTGGC-AGGCCTTC-CCGATAC--AGGGC-TT
AGCCCCGACGCCACCGCGCGCACCGGCACC-AG-G--ATCCGAGCAAGGAGTGCGCCGGCGCCAAGGGCGAG
  2940        2950        2960        2970        2980        2990        3000

400         410         420         430         440         450         460
CAGAGCAAGCAGCTGTGTGCAGGTGTCCTTCAGGAG--GTGGGAGATGTTGGTCTGGACGAATCGCAGACAT
CGGCGCGCGCGGCACCGCGGCGGCCCCCCGAGCGGGGCCCCGGGAGGCGGAGAGCGGG--GAGGAGCCGGCGC
    3010        3020        3030        3040        3050        3060        3070

470         480         490         500         510         520         530
TCTGGTAGGGGCTGGAAGGTACATGAGGTGACAAAATGTATCTCGGTGTTGACGTCCTCCAGAAGCGTTTGC
GGCGGCACCGG--GCCCGGCACA--AGGCG-CAGCCTG-CTCACGAGGCTG------TGGAGAAG-GAGACC
    3080        3090        3100        3110        3120        3130

540         550         560         570         580         590             600
ATCTTAG-ACCCTGCCAC--AGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAGGAAGAG--GCT------CC
A-CGGAGAAGGAGGCCACGGAGAAGGAGGCTGAGATAGTGGAAGC-CGACAAGGAAAAGGAGCTCCGGAACC
      3140        3150        3160        3170        3180        3190        3200

       610         620         630         640         650         660
ACAAGGCCTTGCAGTGCTTCT-CGTCCTGAAGATTGACGGCCACAGTGAC--TGGGTAATCTTTAAGCAGGT
ACCAGCCCCGGGAGCCACACTGTGACCTGGAGA-CCAGTGGGACTGTGACTGTGGGTCCCATGCACACA-CT
     3210        3220        3230        3240        3250        3260        3270

670         680         690         700         710         720         730
GGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGACTGTGGCTGAAGTAACAGTCAG---
GCCCAG-CACCTGTCTCCA---------GAAGGTGGAG--GAACAGCCAGAGGATGCAG--ACAATCAGCGG
      3280                3290        3300        3310        3320

740         750             760         770         780             790
-GTGTCCCCCGCA--GGCA---AGGACTCAGCAGCAGCAACAGCA--GCA------ACAGGGAGGAATTTGG
AACGTCACTCGCATGGGCAGTCAGCCCCCAG-ACCCG-AACACTATTGTACATATCCCAGTGATGCTGACGG
3330        3340        3350        3360        3370        3380        3390

 800         810         820             830         840         850
GCTCCAGGCTGGCGCCAG-CACTGTC-ATCTCTG-----CCGGGGACCCCTCATGCCTGTGACAG----AGC
GCCC---TCTTGGGGAAGCCACGGTCGTTCCCAGTGGTAACGTGGACCTGGAAAGCCAAGCAGAGGGGAAGA
3400        3410        3420        3430        3440        3450        3460

860         870         X
```



```
AGGTCGTCTCGTTCCAGTCGAC
AGGAGG--TGGAAGCGGATGACGTGATGAGGA
3470        3480        3490
```

10. US-08-162-407-1' (1-879)
N60309 Equine IFN-omega-1 from pAH61.

ID N60309 standard; DNA; 2109 BP.
AC N60309;
DT 10-JUN-1991 (first entry)
DE Equine IFN-omega-1 from pAH61.
KW IFN-omega-1; equine; interferon; ss.
OS Equus cabalus.
FH Key Location/Qualifiers
FT CDS 1421..2005
FT /*tag= a
FT /product= IFN-omega-1
FT sig_peptide 1421..1490
FT /*tag= b
FT mat_peptide 1491..2005
FT /*tag= c
PN EP-186098-A.
PD 02-JUL-1986.
PF 17-DEC-1985; 116083.
PR 18-DEC-1984; DE-446122.
PA (BOEH) BOEHRINGER INGELHEIM.
PI Himmler A, Hauptmann R, Hauel N, Adolf G, Swetley P.
DR WPI; 86-170649/27.
DR P-PSDB; P60400.
PT New equine and canine interferon - and recombinant DNA molecules
PT coding for them, and transformed cells.
PS Disclosure; Fig 12; 149pp; German.
CC Microorganisms transformed with the recombinant sequence produce IFN
CC which is useful therapeutically in veterinary medicine.
CC See also N60306-13 and N60938.
SQ Sequence 2109 BP; 609 A; 476 C; 471 G; 553 T;

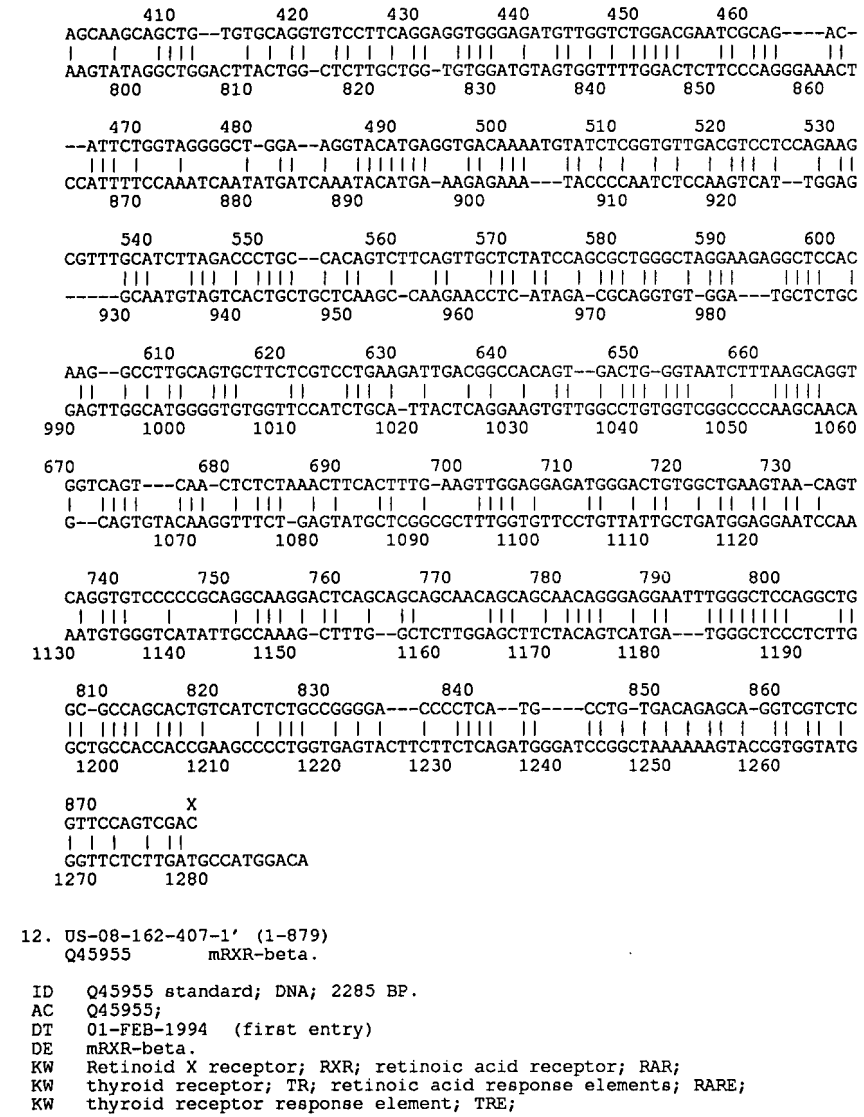
Initial Score = 115 Optimized Score = 340 Significance = 6.77
Residue Identity = 52% Matches = 407 Mismatches = 273
Gaps = 98 Conservative Substitutions = 0

```
          X       10        20        30        40          50
          GTCGACTGGAACGAGACGACCTGCTCCAGACCATTCAGCAAT--CCTGCTCTG-GGCTCCCA
CTCTCAATCTCCCGTCT-CAACCAGAAGAGCTGCGTCTTGGCACTTACCAATGGCTTTCTCAGTGTCTTCC-
    1380 X     1390      1400      1410      1420      1430      1440

60        70          80          90         100         110         120
AGGGCTGAGATTACAGGTG-TAAC--CGAGATAAGAC-CCTGGCTGAGTCAACGA--TGC--ACAAGGCTCG
-CTGATGGCACTGGTGGTGATCTCCTCCAGCCCCGTCTCCTCCATGAG-CTGCGACCTGCCTGCGAGCCTTG
      1450      1460      1470      1480      1490      1500      1510

     130       140                 150       160       170       180
AATCCTAGGGATGGGAGGGG------AGGG----GCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCCCTT
A---CTTGAGAAAGCAGGAGACCCTCAGAGTTCTGCACCAGATGGAGACAATCTCTCCTCCTTCCTGTCTGA
        1520      1530      1540      1550      1560      1570      1580
```

```
      190         200         210         220         230         240         250
TGC-CAGCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAG-CAGCAGCAGCAGGAGCAACA
 || |||   || |  ||||   |  |||| |||| | |  ||  ||||||    |||  ||    || ||
AGCACAGGACAGACTTCAGG-TTCCCCCAGGAGCAGCTGGATGGCAGGCAGTTCCCAGAGGCCCAGGCCACG
      1590        1600        1610        1620        1630        1640        1650

      260         270         280         290         300         310         320
GCTG-CCT-GGGCCGAGGCTCTGGGAGCTCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAGGGT
 ||| |||   | || ||||  | || ||    || |||    | |  |  | | | |   || ||  ||
TCTGTCCTCCAGGAGATGCTCCAGCAGATCGTCAGCCTCTTCCAC-ACAGAGCGC-TCGTCTGCTGCCTGGA
     1660        1670        1680        1690        1700        1710        1720

  330         340         350         360         370         380         390
GGAGGAGTCCGGCTGGCACTGCACCTCCAGGCAC-CGA-GAGAAATTCTGGCAGGCCTTCCCGATACAGGGC
  | || | | |||||  |  | |||| || || | |  || |   |||| || |||   | | ||  |
ACACGACT-CTGCTGGACCGACTCCTCGCGGGACTCCATCAGCA--GCTGGAAGACCT---CAACACCTGCT
      1730        1740        1750        1760        1770        1780

      400         410         420         430         440         450         460
TTCA-GAGCAAGCAGCTGTGTGCAGGTGTCCTTC--AGGAGGTGGGAGATGTTGGTCTGGACGAATCGCAGA
|  | |||||  |||  | | | | |  |||  |   |||  || | |    |  |||| ||    |  |
TGGATGAGCAGACAG--GAGAGGAAGAATCCGCCCTGGGAACTGTG-GGCCCTACACTGGCCGTGAAGAGGT
1790        1800        1810        1820        1830        1840        1850

      470         480         490         500         510         520         530
CATTCTGGTAGGGGCTGGAAGGTACATGA-GGTGACAAAATGTA-TCTCGGTG--TTGACGTCCTCCAGAAG
  ||| || ||   |  |  |||| ||||  | ||  |||| | |  | ||| | |  |    |||| ||
ACTTCAGG-AGAATC-CGTCTGTACCTGACAGAGAAGAAATACAGTGACTGTGCCTGGGAGATTGTCAGA--
1860        1870        1880        1890        1900        1910        1920

        540         550         560         570         580         590
CGTTTGCATC-TTAGACCCTGCCACAGTCTTCAGTTGCTCTATCCAGC--GCTGGGCTAGGAA-GA-GGCT-
 ||   |||| | ||| ||| |     |||||||  |  | | || ||  |   || |||||| || || |
-GTGGACATCATGAGATCCTTC-----TCTTCA--TCAGCAAACCTGCAAGGAAGGTTAGGAATGAAGGATG
      1930        1940        1950        1960        1970        1980

600             610         620         630         640         650         660
-CCACAAGG------CCTTGCAGTGCTTCTCGTCCTGAAGATTGACGGCCACAGTGACTGGGTAATCTTTAAG
   ||  ||      ||||| | || ||||| | |  | | ||  |||||   ||     | |  ||| |
GAGACCTGGGGTCACCTTGAAATGATTCTCCTAC--ACTACTG--GGCCA---TG-----GCACCCTTGCA-
1990        2000        2010        2020            2030            2040

      670         680         690         700         710         720         730
CAGGTGGTCAGTCAACTCTCTAAA--CTTCACTTTGAAGTTGGAGGAGATGGGACTGTGGCTGAAGTAACAG
|  ||  | |||||  | || ||| ||    |||    || |  | ||  |||  |||  |   |||| |
CCTGTCTTTAGTCA--TTTCAAAAGGCTCTTATTTCTGCTTTGGTCATAT---ACT-TTATTGAATTC
2050        2060        2070        2080        2090            2100        X

      740
TCAGGT


11. US-08-162-407-1' (1-879)
    Q03541       Chinese hamster IMPDH.

 ID   Q03541 standard; DNA; 1620 BP.
 AC   Q03541;
 DT   31-AUG-1990  (first entry)
 DE   Chinese hamster IMPDH.
```



```
 KW   Inosine 5'-mono-phosphate dehydrogenase; hepatomas;
 KW   guanosine monophosphate; ss.
 OS   Cricetulus sp.
 PN   WO9001545-A.
 PD   22-FEB-1990.
 PF   02-AUG-1989; U0344.
 PR   12-AUG-1988; US-232302.
 PA   (ARCH-) Arch. Dev. Corp.
 PI   Collart FR, Huberman E;
 DR   WPI; 90-083504/11.
 DR   N-PSDB; Q03541.
 PT   DNA encoding eukaryotic inosine 5'-mono-phosphate dehydrogenase -
 PT   used to detect high levels of enzyme activity of tumour cells,
 PT   esp. hepatomas, and to produce guanosine monophosphate.
 PS   Claim 8; Fig 2; 51pp; English.
 CC   The sequence can be used to quantitatively detect IMPDH encoding
 CC   DNA or RNA by hybridisation, eg in normal and malignant cells,
 CC   esp. hepatomas.
 CC   See also Q03540.
 SQ   Sequence   1620 BP;   423 A;   369 C;   421 G;   406 T;   1 Others;

Initial Score    =    114  Optimized Score  =    387  Significance =  6.70
Residue Identity =    49%  Matches         =    463  Mismatches   =   368
Gaps             =    107  Conservative Substitutions              =     0

             X       10          20          30          40          50
             GTCGACTGGAAC--GAGACG-ACCTGCTCCAGACC--ATTCA-GCAATCCTGCTCTGGGCTC
               | |    ||| |  | || | |  | |  | |  ||  |  || |||   || || ||| ||
   GTAGVCCTTAGCCCCAAGGATCGTGTGAGGGATGTTTTTGAAGCCAAAGCCAGGCATGGCTTCTGTGGTATC
380         390         400         410         420         430         440         450

     60          70          80          90          100         110         120
CCAAGGGCTGAGATTACAGGTGTAACCGAGATAAGACCCTGGCTGAGTCAACGATGC-ACAAGG---CTCGA
|| |   |  |||| |||||     |  | ||   ||  |||| || | || |  | |  |||||    | ||
CCCA--TCACAGAT-ACAGGCCGGATGGGGAGTCGA--CTGG-TG-GGCATCATTTCTTCAAGGGATATTGA
          460         470         480         490         500         510

       130         140         150         160         170         180         190
-ATCCT-AGGGATGGGAGGGGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCCCTTTGCCAGCGA
  | || | ||| || || | |  |     |  || |||| ||    |              ||||  | |
TTTTCTCAAGGA-GGAAGAGCATGACCGTTTCTTGGAGGAGATCATGACAAAGAGGGAAGATTTG---GTGG
 520         530         540         550         560         570         580

   200         210         220         230         240         250         260
AGGCCC---CAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAG-CAGGAGCAACAGCTGCC-
 |||||   |||||  |   | | || |      ||| |  | ||||| |||| |  ||  | || | || ||||
TGGCCCCTGCAGGCATCACTCTGAAGGAGGCAAATGAAATTCTGCAGC-GCAGTAAAAAGGGAAAGTTGCCC
   590         600         610         620         630         640         650

        270         280         290         300         310         320         330
-TGGGCCGAGGCTCTG-GGAGCTCCGTGGCTTCTAGGGCTATGGGACTCCTTGGGGGCAGCAGGGTGGAGGA
 | |     |  ||  |||||  || ||    |  |||  |  |  |||  |  | || |  || ||  |
ATTGTGAATGAAAATGATGAGCT-GGTAGCCATCATTGCTCGGACAGACCT--GAAGAAGAATCGT-GATTA
  660         670         680         690         700         710         720

          340         350         360         370         380         390         400
GTC--CGGCTGGCACTG-CACCTCCAGGCACCGAGAGAAATTCTGGCAGGCCTTCCCGATACAGGGCTTCAG
  |   ||||  ||  |   ||  | |||  |  |  |  |  ||||| | ||     |   | |  | |
CCCATTGGCTTCCAAAGATGCCAAAAAGCA--GCTACTATGTGGGGCAGCCATTGGTACTCATGAGGATGAC
    730         740         750         760         770         780         790
```

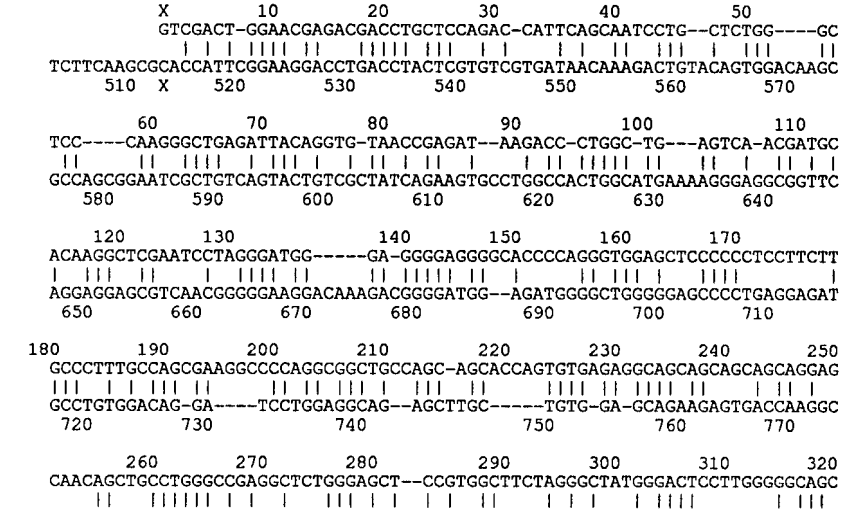


12. US-08-162-407-1' (1-879)
Q45955 mRXR-beta.

ID Q45955 standard; DNA; 2285 BP.
AC Q45955;
DT 01-FEB-1994 (first entry)
DE mRXR-beta.
KW Retinoid X receptor; RXR; retinoic acid receptor; RAR;
KW thyroid receptor; TR; retinoic acid response elements; RARE;
KW thyroid receptor response element; TRE;

KW retinoid X receptor response element; RXRE;
KW subunit; heterodimer; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 76..1422
FT /*tag= a
FT /label= RXR-beta
PN WO9315216-A.
PD 05-AUG-1993.
PF 25-JAN-1993; U00639.
PR 24-JAN-1992; US-825667.
PA (CNRS) CENT NAT RECH SCI.
PA (INRM) INST NAT SANTE & RECH MEDICALE.
PA (SQUI) SQUIBB & SONS INC E R.
PA (UYST-) UNIV PASTEUR STRASBOURG LOUIS.
PI Chambon P, Kastner P, Leid M;
DR WPI; 93-258691/32.
DR P-PSDB; R39467.
PT Heterodimer comprising retinoic acid or thyroid receptor and
PT retinoid X receptor - useful in diagnosis and for generation of
PT antibodies
PS Claim 26; Page 84-86; 128pp; English.
CC Two different types of nuclear receptors, RAR and TR dimerise with
CC RXR to form a heterodimer which is capable of binding to RARE, TRE
CC or RXRE at physiological conditions. Sequences encoding 1 isoform
CC of mouse RXR-beta, 3 isoforms of human RXR-beta, 1 isoform of mouse
CC RXR-alpha and 1 isoform of mouse RXR-gamma are provided (Q45955-60).
SQ Sequence 2285 BP; 447 A; 665 C; 681 G; 492 T;

Initial Score = 113 Optimized Score = 386 Significance = 6.63
Residue Identity = 49% Matches = 471 Mismatches = 360
Gaps = 124 Conservative Substitutions = 0



```
GTTGAGGGTCCTGGGGCCA--CCGGGGGTGGTGGCAGCAGCCCAAATGACCCAGTGACTAACATCTGCCAGG
   780        790         800        810        820        830        840

        330        340         350         360        370        380
AGGGTGGAGGAGTCCGGCTG-GCACTGCACCTCCAG-GCACCGAGAGAAATTCTGGCAGGCCT--TCCCGA-
  | | ||   |   | ||||  |||  | | |  || |  |  |||| |  ||  |||   ||  |||| |
CAGCT-GACAA--ACAGCTGTTCAC-ACTCGTTGAGTGGGCAAAGAGGA--TCCCGCACTTCTCCTCCCTAC
    850         860        870        880         890        900        910

390        400        410        420        430        440        450
TACAGGGC-TTCAGAGCAAGCAGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGGTCTGGACGAAT
  | || |  ||||  ||  | ||||  | |||||   |    | |||  |    ||   |  ||   |  ||
CTCTGGACGATCAGGTCATACTGCTGCGGGCAGG---CTGGAACGAGCT-CCTCATTGCGTCCTTCTCCCAT
       920        930        940         950        960        970

460        470        480        490         500        510        520
CGCAGACATTCTGGTAGGGGCTGGAAGGTACATG---AGGTGACAAAATGTATCTCGGTGTTGACGTCC--T
||    ||||   ||  | | ||| |    | ||   | | | |   ||||  | | |       | ||  |
CG-GTCCATTGATGTCCGAGATGGCATCCTCCTGGCCACGGGTCTTCATGT-GCACAGAAACTCAGCCCATT
980        990       1000       1010       1020       1030       1040

   530          540          550        560        570        580
CCAGAAGCGTTTG---CATCTTAGACC-----CTGCCACAGTCTTCAGT-TGCTCTAT-CCAGCGCTGGGCT
||  | ||||  |   ||||||| || |     ||| || || ||  ||| | |   || |  |   || |
CCGCAGGCGTGGGAGCCATCTTTGATCGGGTGCTGACAGAG-CT--AGTGTCCAAAATGCGTGACATGAGGA
1050       1060       1070       1080       1090       1100       1110

590        600         610        620        630        640        650
AGGAAGAGGCTCCACAAGGC--CTTGCAGTGCTTC-TCGTCCTGAAGATTGACGGCCACAGTGACTGGGTAA
 |||  || |    |  |||  | ||| |    || || |  | ||    ||| |||| || | ||    ||
TGGACAAGACAGAGCTTGGCTGCCTGCGGGCAATCATCATGTTTAATCCAGAC-GCCA-AGGGCCTCTCCAA
1120       1130       1140       1150       1160       1170       1180

 660        670              680        690        700        710        720
TCTTTAAGCAGGTGG---TCAGTC----AACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGACTGTG
 | |  || ||||||   ||  ||      |   | ||  | |||||    |  |   | | |   ||  ||
CCCTGGAG-AGGTGGAGATCCTTCGGGAGAAGGTGTACGCCTCACTGGAGACCTATTGCA-AGCAGA-AGTA
1190       1200       1210       1220       1230       1240       1250

      730        740        750        760        770        780
GCTGAAGTAACA-GTCAGGTGTCCCCCGCAG--GCAAGGACTCAGCAGCAGCAACAGCAGCAACAG--GGAG
 |   || | || | | ||| | ||  || |  |  | | ||   | ||  |  | ||  || | |    |
CCCTGAGCAGCAGGGCCGGTTTGCCAAGCTGCTGTTACGTCT-TCCTGC--CCTCCGCTCCATCGGCCTCAA
1260       1270       1280       1290       1300       1310       1320

790        800        810        820        830        840        850
GAATTTGG-GCTCCAG-GCTGGCGCCAGCACTGTCATCTCTGCCGGGGACCCC--TCATGCCT--GTGACAG
|  | ||| || || |  ||    | |||    ||||   || ||    ||||  | |  |||   | |  |
GTGTCTGGAGCACCTGTTCTTCTTCAAGC----TCAT---TGGCGACACCCCCATTGACACCTTCCTCATGG
 1330       1340       1350             1360       1370       1380       1390

   860          870         X
AGCAGGT--CGTCT-CGTTCCAG-TCGAC
||  | |   | || |   |||| | | |
AGATGCTTGAGGCTCCCCACCAGCTAGCCTGAGCCCAGA
      1400       1410       1420
```

13. US-08-162-407-1' (1-879)
    Q43031 Collagen-like polymer DCP2-C2(A12)C2 coding sequen


```
ID   Q43031 standard; DNA; 432 BP.
AC   Q43031;
DT   07-SEP-1993  (first entry)
DE   Collagen-like polymer DCP2-C2(A12)C2 coding sequence.
KW   Recombinant; collagen-like polymer; CLP; tripeptide; helix; membrane;
KW   fibre; film; coating; triad sequence; collagen; mammalian; moulding;
KW   hydrogel; interchain linkage; colloid suspension; DCP; antibody; ds.
OS   Synthetic.
PN   WO9310154-A.
PD   27-MAY-1993.
PF   04-NOV-1992; U09485.
PR   12-NOV-1991; US-791960.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J,  Ferrari FA;
DR   WPI; 93-182496/22.
DR   P-PSDB; R37737.
PT   High mol. wt. collagen-like protein polymers - capable of being
PT   produced in unicellular microorganisms
PS   Disclosure; Page 38; 82pp; English.
CC   The sequences given in Q43030-34 encode examples of recombinantly
CC   produced DCP collagen-like polymers (CLPs) which consist of repeated
CC   tripeptide sequences selected from a wide range of GXY sequences,
CC   where X and Y can be any amino acid.  These sequences can be cloned
CC   into plasmids and used to transform E. coli to produce the DCP
CC   proteins.  DCP peptides comprise repeated units of; A = GAPGPAGPP,
CC   B = GSRGDPGPP and/or C = GAHGPAGPK.  These polymers may be used to
CC   raise anti-DCP antibodies in rabbits.  These polymers have molecular
CC   weights of >30 kD and are able to form helices due to interchain
CC   linkages.  These polymers pref. contain a proportion of tripeptide
CC   triad sequences found in natural collagens, pref. mammalian collagens.
CC   The CLPs impart unique characteristics to materials such as fibres,
CC   membranes, films, coatings, hydrogels, colloid suspensions and moulded
CC   articles.
SQ   Sequence   432 BP;    54 A;    162 C;    165 G;    51 T;

Initial Score    =    112  Optimized Score  =    204  Significance =  6.56
Residue Identity =    51%  Matches          =    239  Mismatches   =   177
Gaps             =     48  Conservative Substitutions               =     0

      190    X  200        210        220        230        240        250
TTGCCAGCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCAACAG
             |  | ||  |||| |||||||  ||   | | |  |||   ||  |||||||  | | ||
          GGTGCTCA--CGGC--CCAGCAGGTCCGAAGGGCG-CGCA-TGGC-CCAGCAGGCCCGAAAG
          X          10         20         30         40         50

      260        270        280        290        300        310
CT--GCCTGGGCCG--AGG--CTCTGGGAGCTCCGTGGCTTCTAGG---GCTATGGGACTCCTTGGGGGCAG
 |  |||||| |||   ||  | | ||| |||||| | | |  |||   || | ||  | |||  ||  | |
GTGCGCCTGGACCGGCTGGTCCACCGGGTGCTCCGGGACCTGCAGGCCCGCCA-GGTGCGCCT--GGACCGG
   60         70         80         90        100        110        120

  320        330        340        350        360        370        380        390
CAGGGTGGAGGAGTCCGGCTGGCACTGCACCTCCAGGCACCGAGA-GAAATTCTGGCAGGCCTTCCCGATAC
|  |||  |      |||| ||   | | |||| ||||| |||  | |     ||||   | ||   | |
C-TGGTCCA------CCGGGTGCTCCGGGACCTGCAGGC-CCGCCAGGTGCGCCTGGACCGGCTGGTCCACCG
    130            140        150        160        170        180

        400        410        420        430        440        450
AGGGCTTCAGAGCAAGCAGCTGTGTGCAGGTG-TCCTTCAGGAGGTGG--GAGATGTTGGTCTGGACGAATC
```

```
  | ||| | |  |  ||||    |  ||||||  |||  |  | |||  |  | || || ||  ||
  GGTGCTCCGGGACCTGCAGGCCCG-CCAGGTGCGCCTGGACCGGCTGGTCCACCGGGTGCTCCGG--GACCT
190       200       210       220       230       240       250

460       470       480       490       500       510       520       530
  GCAGACATTCTGGTAGGGGCTGGAAGGTACATGAGGTGACAAAATGTATCTCGGTGTTG-ACGTCCTCCAGA
  |||| |   |  | | ||||  |    || |  ||    ||  ||  ||   || | | || |||| |
  GCAGGCCCGCCAGGTGCGCCTGGACCG--GCTG-GTCCACCGGGTG-CTCCGGGACCTGCAGGCCCGCCAGG
260       270       280       290       300       310       320

        540       550       560       570       580       590       600
  AGCGTTTGCATCTTAGACCCTGCCACAGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAG-GAAGA-GGCTCC
   |||  || | |   | |    |||| |  | |    |  |  | ||| ||  | |  | |   | ||| |
  TGCGCCTGGACC---GGCTGGTCCACCGGGTGCTCCGGGAC-CTGCAG-GCCCGCCAGGTGCCCATGGCCCA
   330           340       350       360       370       380       390

        610       620       630       640 X     650
  ACAAGGCCTTGCAGTGCTTCTCGTCCTGAAGATTGACGGCCACAGTGACTGG
   || | ||     | |||    |||| | ||   | | |  |
  GCAGGACCGAAAGGAGCTCACGGTCCGGCAG---GTCCGAAA
      400       410       420       430 X
```

14. US-08-162-407-1' (1-879)
    Q43032 Collagen-like polymer DCP3-C2(AB12)C2 coding seque

```
ID   Q43032 standard; DNA; 756 BP.
AC   Q43032;
DT   07-SEP-1993  (first entry)
DE   Collagen-like polymer DCP3-C2(AB12)C2 coding sequence.
KW   Recombinant; collagen-like polymer; CLP; tripeptide; helix; membrane;
KW   fibre; film; coating; triad sequence; collagen; mammalian; moulding;
KW   hydrogel; interchain linkage; colloid suspension; DCP; antibody; ds.
OS   Synthetic.
PN   WO9310154-A.
PD   27-MAY-1993.
PF   04-NOV-1992; U09485.
PR   12-NOV-1991; US-791960.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J,  Ferrari FA;
DR   WPI; 93-182496/22.
DR   P-PSDB; R37738.
PT   High mol. wt. collagen-like protein polymers - capable of being
PT   produced in unicellular microorganisms
PS   Disclosure; Page 39; 82pp; English.
CC   The sequences given in Q43030-34 encode examples of recombinantly
CC   produced DCP collagen-like polymers (CLPs) which consist of repeated
CC   tripeptide sequences selected from a wide range of GXY sequences,
CC   where X and Y can be any amino acid.  These sequences can be cloned
CC   into plasmids and used to transform E. coli to produce the DCP
CC   proteins.  DCP peptides comprise repeated units of; A = GAPGPAGPP,
CC   B = GSRGDPGPP and/or C = GAHGPAGPK.  These polymers may be used to
CC   raise anti-DCP antibodies in rabbits.  These polymers have molecular
CC   weights of >30 kD and are able to form helices due to interchain
CC   linkages.  These polymers pref. contain a proportion of tripeptide
CC   triad sequences found in natural collagens, pref. mammalian collagens.
CC   The CLPs impart unique characteristics to materials such as fibres,
CC   membranes, films, coatings, hydrogels, colloid suspensions and moulded
CC   articles.
SQ   Sequence   756 BP;     84 A;     288 C;     291 G;     93 T;
```



```
Initial Score    =    112  Optimized Score  =    315  Significance =  6.56
Residue Identity =    51%  Matches          =    370  Mismatches   =   272
Gaps             =     74  Conservative Substitutions               =     0

      190    X  200       210       220       230       240       250
   TTGCCAGCGAAGGCCCCAGGCGGCTGCCAGCAGCACCAGTGTGAGAGGCAGCAGCAGCAGGAGCAACAG
                |  | ||  ||||  |||||||  ||   | | |  |||   ||  |||||||  | | ||
              GGTGCTCA--CGGC--CCAGCAGGTCCGAAGGGCG-CGCA-TGGC-CCAGCAGGCCCGAAAG
              X          10          20          30          40          50

      260       270       280       290       300       310
   CT--GCCTGGGCCG--AGG--CTCTGGGAGCTCCGTGGCTTCTAGG---GCTATGGGACTCCTTGGGGGCAG
    |  |||||| |||   ||  | | ||| |||||| | | |  |||   || | ||  | |||  ||  | |
   GTGCGCCTGGACCGGCTGGTCCACCGGGTGCTCCGGGACCTGCAGGCCCGCCA-GGTGCGCCT--GGACCGG
      60        70        80        90        100       110       120

 320       330       340       350       360       370       380       390
   CAGGGTGGAGGAGTCCGGCTGGCACTGCACCTCCAGGCACCGAGA-GAAATTCTGGCAGGCCTTCCCGATAC
   |  |||  |     |||| ||   | | |||| |||||  |||  | |     ||||   | ||   | |
   C-TGGTCCA-----CCGGGTGCTCCGGGACCTGCAGGC-CCGCCAGGTGCGCCTGGACCGGCTGGTCCACCG
       130           140       150       160       170       180

         400       410       420       430       440       450
   AGGGCTTCAGAGCAAGCAGCTGTGTGCAGGTG-TCCTTCAGGAGGTGG--GAGATGTTGGTCTGGACGAATC
    | ||| | |  |  ||||    |  ||||||  |||  |   | |||   |   | || || ||  ||
   GGTGCTCCGGGACCTGCAGGCCCG-CCAGGTGCGCCTGGACCGGCTGGTCCACCGGGTGCTCCGG--GACCT
 190       200       210       220       230       240       250

 460       470       480       490       500       510       520       530
   GCAGACATTCTGGTAGGGGCTGGAAGGTACATGAGGTGACAAAATGTATCTCGGTGTTG-ACGTCCTCCAGA
   |||| |   |  | | ||||  |    || |  ||    ||  ||  ||   || | | || |||| |
   GCAGGCCCGCCAGGTGCGCCTGGACCG--GCTG-GTCCACCGGGTG-CTCCGGGACCTGCAGGCCCGCCAGG
 260       270       280       290       300       310       320

         540       550       560       570       580       590
   AGCGTTTGCATCTTAGACCCTGCCACAG---TCTTCAGTTGCT-CTATCCAGC--GCTGGGCTAGGAAGAGG
    |||  || | |   | |    |||| |   || | |   || |   || ||  | || ||  || | ||
   TGCGCCTGGACC---GGCTGGTCCACCGGGTGCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGG-ACCGG
    330           340       350       360       370       380       390

    600       610       620       630       640       650       660
   CTCCACAAGGCCTTGCAGTGCTTCTCGTCCTGAAGATTGACGGCCA-CAGTGACTGGGTAATCT--TTAAGC
   ||   | |  ||     ||||| |  | |||| ||   | | ||||   | | ||||    ||  |  | |
   CTGGTCCA--CC---GGGTGCTCCGGGACCTGCAG---GCCCGCCAGGTGCGCCTGGACCGGCTGGTCCACC
      400           410       420       430       440       450

     670       680       690       700       710       720       730
   AGGTGGTCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGG--GACTGTG-GCTGAAGTAACAG
    |||| || |  | ||  |    |  |    ||    ||||   | |||   || | | |||    | | | |
   GGGTGCTCCGGGACCT-GCAGGCCCGCCAGGTGCGCCTGGACCGGCTGGTCCACCGGGTGCTCCGGGACCTG
 460       470       480       490       500       510       520

      740       750       760       770       780       790       800
   TCAGGTGTCCCCCGCAGGCAAGGACTCAGCAGCAGCAACAGCAGCAACAGGGAGGAATTTGGGCTCCAGGCT
    ||||    ||  |  ||  |||| | || |   || | |  ||  | ||||       |  | ||||| |
   -CAGGCCCGCCAGGTGCGCCTGGAC-CGGCTGGTCCACCGGGTGC-TCCGGGACCTGCAGGCCCGCCAGG-T
 530       540       550       560       570       580       590

   810       820       830       840       850       860       870
```

```
GGCGCCAGCACTGTCATCTCTGCCGGGGACCCCTCATGCCTG-TGACAGAGCAGGT-CGTCTCGTTC--CAG
| |||| | || | |   ||  |||||  | ||     ||||  | |    ||||| || || |  |  | |
G-CGCCTGGACCGGCTGGTCCACCGGGTGCTCC-GGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTG
 600       610       620       630       640       650       660

     X
-TCGAC
 || ||
GTCCACCGGGTGCTCC
670  X    680
```

15. US-08-162-407-1' (1-879)
    Q37818 Sequence encoding the alpha 1B-2 human calcium cha

```
ID   Q37818 standard; cDNA; 7175 BP.
AC   Q37818;
DT   30-JUN-1993  (first entry)
DE   Sequence encoding the alpha 1B-2 human calcium channel subunit.
KW   Human calcium channel subunit; diagnosis; agonist; antagonist;
KW   Lambert Eaton syndrome; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             144..6857
FT   /*tag=  a
PN   WO9304083-A.
PD   04-MAR-1993.
PF   14-AUG-1992; U06903.
PR   15-AUG-1991; US-745206.
PR   10-APR-1992; US-868354.
PA   (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI   Brenner R,  Ellis SB,  Feldman DH,  Harpold MM,  Mccue AF,
PI   Williams ME;
DR   WPI; 93-093936/11.
DR   P-PSDB; R33550.
PT   DNA encoding specific human calcium channel sub-units - used for
PT   identifying calcium channel agonists and antagonists and
PT   diagnosing Lambert Eaton syndrome
PS   Disclosure; Page 120-128; 150pp; English.
CC   DNA encoding the alpha 1B subunit was isolated by screening a
CC   human basal ganglia cDNA library with  fragments of the rabbit
CC   skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.
CC   A portion of one of the positive clones was used to screen an IMR32
CC   cell cDNA library. Clones that hybridized to the basal ganglia
CC   DNA prove were used to further screen an IMR32 cell cDNA library
CC   to identify overlapping clones that in turn were used to screen a
CC   human hippocampus cDNA library. In this way, a sufficient series of
CC   clones to span nearly the entire length of the nucleotide sequence
CC   encoding the human alpha 1B subunit was obtained. PCR amplification
CC   of specific regions of the IMR32 cell alpha 1B mRNA yielded
CC   additonal segments of the alpha 1B coding sequence. A full-length
CC   alpha 1B DNA clone was constructed by ligating portions of the
CC   partial cDNA clones (see Q37817, Q37818). Alpha 1B-1 and alpha
CC   1B-2 are derived by alternative splicing of the alpha 1B subunit
CC   transcript.
SQ   Sequence   7175 BP;    1415 A;    2204 C;    2162 G;    1394 T;
```

Initial Score	= 112	Optimized Score	= 404	Significance	= 6.56
Residue Identity	= 51%	Matches	= 488	Mismatches	= 336
Gaps	= 122	Conservative Substitutions			= 0


```
              X         10        20        30        40        50
             GTCGA-CTGGAACGAGACGACCTGCTCCAGACCAT-TCAG-CA-ATCCTGCTCTG-GGCTCC
             || || |||| || |||| |  ||    | || |   || || | || |  ||| |||| |
   GCCGCTGGTGGTGGAGCTGGGCCGCGACGGCGCGCGGGGGCCCGTGGGAGGCAAAGCCCGACCTGAGGCTGC
2590       2600      2610      2620      2630      2640      2650

  60        70        80         90         100       110       120
  CAAGGGCTGAGATTACAGGTGTAACC-GAGATAAG--ACC-CTGGCTGAGTCAACGATGCACAAG-GCTCGA
   ||| |   ||  | ||  | ||   |   ||  ||| | |||   | | || | | |||||  | |
   GGAGGCCCCCGAGGGC--GTCGACCCTCCGCGCAGGCACCACCGGCACCG-CGAC-AAGGACAAGACCCCCG
2660       2670      2680      2690      2700      2710      2720

     130       140       150       160       170       180       190
  ATCCTAGGGA-TGGGA--GGGGAGGGGCACCCCAGGGTGGAGCTCCCCCCTCCTTCTTGCCCTTTGCCAGCG
     | ||||   |||  | | || ||| ||| | || |||| |      ||    |||||      ||||
  CGGCGGGGGACCAGGACCGAGCAGAGGC-CCCGAAGGCGGAGAGCGGGGAGCC-CGGTGCCCGGGAGGAGCG
2730       2740      2750      2760      2770      2780      2790

         200       210       220       230       240       250       260
  ---AAGGCCCCAGGCGGCTGCCAGCAGC-ACCAGTGTGAGAGGCAGCAGCAGCAGCAGGAGCAACAGCTGCC
      |||| ||  | || |||| |||| |  ||   | | |||  | | ||  || ||||| |  || |||
  GCCGCGGCCGCA--CCGCAGCCA-CAGCAAGGAGGCCGCGGGGCCCCCGGAGGCGC-GGAGCGAGCGCGGCC
2800       2810      2820      2830      2840      2850      2860

        270       280       290       300       310       320
  TGGGCCGAGGCTCTGGGAGC-TCCG-TGGCTTCTAGGGCTATGGGACTCCTTG--GGGGCAGCAGGGTGGAG
    |||| |||| | | | ||  |||  |||  |   |||   | | |||| |  || || ||   | | |
  GAGGCCCAGGCCCCGAGGGCGGCCGGCGGCACCACCGGC---GCGGCTCCCCGCAGGAGCCGC-CCGAGCGG
   2870      2880      2890      2900      2910      2920      2930

330        340       350       360       370       380       390
  GAGTCCGGCTGGCACTGCACCTCCAGGCACCGAGAGAAATTCTGGC-AGGCCTTC-CCGATAC--AGGGCTT
  ||| || |  | ||| || |   | |||||| || |  || |  || |||  | | |||   |  |||||
  GAGCCCCGACGCCACCGCGCGCACCGGCACC-AG-G--ATCCGAGCAAGGAGTGCGCCGGCGCCAAGGGC--
     2940      2950      2960      2970      2980      2990

  400       410       420       430       440       450       460
  CAGAGCAAGCAGCTGTGTGCAGGTGTCCTTCAGGAGGTGGGAGATGTTGGTCTGGACGAATCGCAGACATTC
  ||| |   |  || |      || | ||  |  |||  |||     |   | | ||| ||  ||  ||    |
  CAGCGGCGGGCGCGGCACCGCGGCGGCC--CCCGAGCGGGGCCCCGGGAGGC-GGA-GA-GCGGGGAGGAGC
3000       3010      3020      3030      3040      3050      3060

470        480       490       500       510       520       530       540
  TGGTAGGGGCTGGAAGGTACATGAGGTGACAAAATGTATCTCGGTGTTGACGTCCTCCAGAAGCGTTTGCAT
   ||  | ||| | |  |  |     |  ||||   | | |  | |    || |   |  ||||| |    ||
  CGG-CGCGGCGGGCACCGGGC---CCGGCACAAGGCGCAGCCTGCTCACGAGGCTGTGGAGAAG-GAGACCA-
   3070      3080      3090      3100      3110      3120      3130

        550       560       570       580       590               600
  CTTAG-ACCCTGCCAC--AGTCTTCAGTTGCTCTATCCAGCGCTGGGCTAGGAAGAG--GCT------CCAC
  |  || |    |||||  ||       | ||   ||       ||  | | |||||  ||  |||      ||||
  CGGAGAAGGAGGCCACGGAGAAGGAGGCTGAGATAGTGGAAGC-CGACAAGGAAAAGGAGCTCCGGAACCAC
     3140      3150      3160      3170      3180      3190      3200

      610       620       630       640       650       660       670
  AAGGCCTTGCAGTGCTTCT-CGTCCTGAAGATTGACGGCCACAGTGAC--TGGGTAATCTTTAAGCAGGTGG
   || ||  | ||      ||  | |||| |||   |  |  || |||||  |||||     |  |  ||  ||
  CAGCCCCGGGAGCCACACTGTGACCTGGAGA-CCAGTGGGACTGTGACTGTGGGTCCCATGCACACA-CTGC
     3210      3220      3230      3240      3250      3260      3270
```

```
       680       690       700       710       720       730
  TCAGTCAACTCTCTAAACTTCACTTTGAAGTTGGAGGAGATGGGACTGTGGCTGAAGTAACAGTCAG----G
   ||| || || ||| |          |||| |||||  ||   | | | || || ||  ||| ||||
  CCAG-CACCTGTCTCCA---------GAAGGTGGAG--GAACAGCCAGAGGATGCAG--ACAATCAGCGGAA
     3280              3290        3300      3310        3320      3330

740       750            760       770         780            790
  TGTCCCCCGCA--GGCA---AGGACTCAGCAGCAGCAACAGCA--GCA------ACAGGGAGGAATTTGGGC
   ||| | ||||  ||||   ||  | ||| | | | ||||  |  | |       ||| || |     ||||
  CGTCACTCGCATGGGCAGTCAGCCCCCAG-ACCCG-AACACTATTGTACATATCCCAGTGATGCTGACGGGC
     3340      3350      3360        3370      3380      3390      3400

800       810        820            830       840       850
  TCCAGGCTGGCGCCAG-CACTGTC-ATCTCTG-----CCGGGGACCCCTCATGCCTGTGACAG----AGCAG
   |    || | |  || ||| |||  || | |       || |||||    | |||    | ||    || ||
  CC---TCTTGGGGAAGCCACGGTCGTTCCCAGTGGTAACGTGGACCTGGAAAGCCAAGCAGAGGGGAAGAAG
        3410      3420      3430      3440      3450      3460      3470

860       870       X
  GTCGTCTCGTTCCAGTCGAC
  |  |  | |   | |  |||
  GAGG--TGGAAGCGGATGACGTGATGAGGA
        3480      3490
```

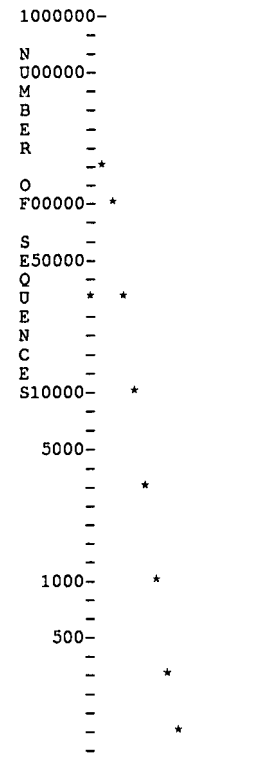
NeWSprinter20

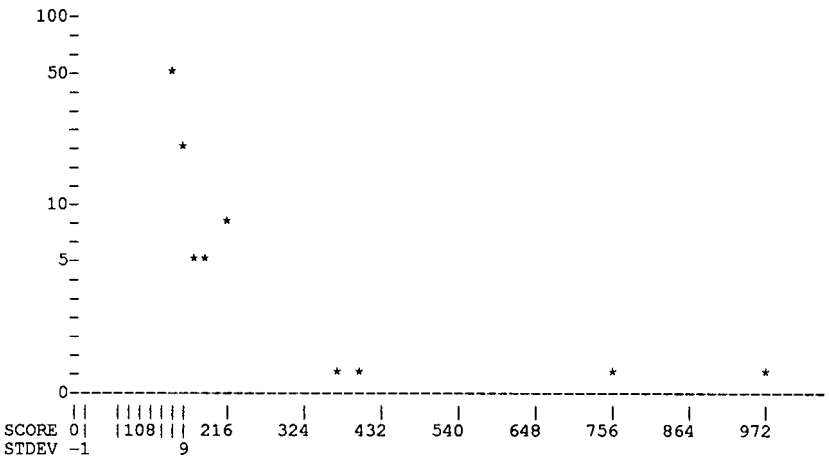
Thu Apr 06 10:12:42 1995

NeWSprint 2.5 Rev B
Openwin library 3
NeWSprint interpreter 210.0

NeWSprint 2.5

```
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq5.res made by  on Wed 5 Apr 95 18:23:51-PDT.


Query sequence being compared:US-08-162-407-5 (1-988)
Number of sequences searched:                313646
Number of scores above cutoff:               3864

     Results of the initial comparison of US-08-162-407-5 (1-988) with:
   Data bank : EMBL-NEW 1, all entries
   Data bank : GenBank 86, all entries
   Data bank : GenBank-NEW 1, all entries
   Data bank : UEMBL 41_86, all entries
```





PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	31	30	13.95

Times:	CPU	Total Elapsed
	00:42:44.01	00:42:48.00

```
Number of residues:              288682119
Number of sequences searched:       313646
Number of scores above cutoff:        3864

Cut-off raised to 22.
Cut-off raised to 27.
Cut-off raised to 30.
Cut-off raised to 33.
Cut-off raised to 36.
Cut-off raised to 39.
Cut-off raised to 42.
Cut-off raised to 45.
Cut-off raised to 49.
Cut-off raised to 52.
```

Cut-off raised to 56.
Cut-off raised to 59.
Cut-off raised to 62.
Cut-off raised to 66.
Cut-off raised to 70.
Cut-off raised to 74.
Cut-off raised to 78.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

	Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
		**** 67 standard deviations above mean ****					
1.	HSU03858	Human flt3 ligand mRNA, compl	1080	972	979	67.45	0
		**** 51 standard deviations above mean ****					
2.	HSU04806	Human FLT3/FLK2 ligand mRNA,	859	747	767	51.32	0
		**** 26 standard deviations above mean ****					
3.	MMU04807	Mus musculus FLT3/FLK2 ligand	1152	400	611	26.45	0
		**** 24 standard deviations above mean ****					
4.	MUSLIGAND	Mouse flt3 ligand mRNA, compl	829	378	580	24.87	0
		**** 13 standard deviations above mean ****					
5.	S65267	drebrin A {alternatively spli	1919	221	404	13.62	0
6.	CHKDEBPDA	Chicken drebrin mRNA.	2102	221	448	13.62	0
7.	HSELASRNA	H.sapiens RNA for dermal fibr	2241	218	405	13.40	0
8.	HSELASRNA	H.sapiens RNA for dermal fibr	2241	218	405	13.40	0
9.	HSELASRNA	H.sapiens RNA for dermal fibr	2241	218	405	13.40	0
10.	HUMELASF	Human elastin mRNA, complete	2242	218	405	13.40	0
11.	HSELASF	Human elastin mRNA, complete	2242	218	405	13.40	0
		**** 11 standard deviations above mean ****					
12.	HSOXYTOC	H.sapiens gene for oxytocin r	3617	193	461	11.61	0
13.	HSOXYTOC	H.sapiens gene for oxytocin r	3617	193	461	11.61	0
14.	HSMRNAOXY	H.sapiens mRNA for oxytocin r	4103	193	456	11.61	0
15.	TCU17917	Thermomonospora curvata alpha	2093	185	450	11.04	0
16.	TC17917	Thermomonospora curvata alpha	2093	185	450	11.04	0
		**** 10 standard deviations above mean ****					
17.	GGGFBPA	G.gallus gammaFBP-A mRNA.	3597	178	454	10.54	0
18.	STMPGLYZ	Streptomyces coelicolor bacte	9461	176	419	10.39	0
19.	SCPGLYZ	Streptomyces coelicolor bacte	9461	176	419	10.39	0
20.	DMHSP82	Drosophila melanogaster gene	5024	173	437	10.18	0
		**** 9 standard deviations above mean ****					
21.	PAXCPSP	P.aeruginosa genes xcpR, xcpS	5940	168	452	9.82	0
22.	AZSNIFA	A.lipoferum gene for NifA pro	2820	167	454	9.75	0
23.	SGSTR	Streptomyces griseus genes st	10653	166	457	9.68	0
24.	PMPEMA62A	Peromyscus maniculatus clone	1573	165	403	9.60	0
25.	MMRNAMHA	M.musculus mRNA for mouse hai	1591	164	436	9.53	0
26.	RABBCCBIII	Rabbit mRNA for rabbit brain	7713	163	437	9.46	0
27.	MZEKINAA	Z.mays protein cdc2 kinase mR	1325	162	436	9.39	0
28.	GGGFBPC	G.gallus gammaFBP-C mRNA.	3411	160	454	9.25	0
29.	GGGFBPB	G.gallus gammaFBP-B mRNA.	3551	160	454	9.25	0
30.	S45791	band 3-related protein=renal	3964	160	451	9.25	0
31.	BOVA1ADRE	Cow A1 adenosine receptor mRN	1977	159	444	9.17	0
32.	HSU11690	Human faciogenital dysplasia	4266	158	452	9.10	0
33.	HS11690	Human faciogenital dysplasia	4266	158	452	9.10	0

34.	HUMOP2A	Homo sapiens osteogenic prote	1842	157	438	9.03	0
35.	HUMOP2A	Homo sapiens osteogenic prote	1842	157	438	9.03	0
		**** 8 standard deviations above mean ****					
36.	SCU03771	Streptomyces coelicolor A3(2)	6630	156	459	8.96	0
37.	HUMPDEAA	Human cyclic AMP (cAMP)-speci	2619	155	456	8.89	0
38.	HSGAA01	Human acid alpha-glucosidase	3441	155	452	8.89	0
39.	PIGI2AA	Porcine G protein gene, exon	983	152	388	8.67	0
40.	OAMGF	O.aries MGF mRNA for mammary	2820	150	448	8.53	0
41.	OAMGF	O.aries MGF mRNA for mammary	2820	150	448	8.53	0
42.	RATBCCA1S	Rat calcium channel alpha-1 s	6639	150	438	8.53	0
43.	AVIHUPZ	Azotobacter chroococcum hydro	5400	148	456	8.39	0
44.	RNDES	R.norvegicus gene for desmin.	2285	147	435	8.31	0
45.	ECPIFC	E. coli miniF plasmid gene pi	1710	146	439	8.24	0

1. US-08-162-407-5 (1-988)
HSU03858 Human flt3 ligand mRNA, complete cds.

```
LOCUS       HSU03858     1080 bp    mRNA            PRI       19-JUL-1994
DEFINITION  Human flt3 ligand mRNA, complete cds.
ACCESSION   U03858
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Lyman,S.D., James,L., Johnson,L., Brasel,K., de Vries,P.,
            Escobar,S.S., Downey,H., Splett,R.R., Beckmann,M.P. and
            McKenna,H.J.
  TITLE     Cloning of the human homologue of the murine flt3 ligand: a growth
            factor for early hematopoietic progenitor cells
  JOURNAL   Blood 83, 2795-2801 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Lyman,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-1993) Stewart D. Lyman, Immunex Research and
            Development Corporation, 51, University St., Seattle, WA 98101, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 494978
FEATURES             Location/Qualifiers
     source          1..1080
                     /clone_lib="expression, cDNA"
                     /organism="Homo sapiens"
                     /cell_line="clone 22 (T cell)"
     5'UTR           1..83
     CDS             84..791
                     /standard_name="FMS-like tyrosine kinase-3 ligand"
                     /note="ligand for the flt3/flk-2 tyrosine kinase receptor;
                      NCBI gi: 494979"
                     /codon_start=1
                     /function="stimulates proliferation of early hematopoietic
                     cells"
                     /product="flt3 ligand"
                     /translation="MTVLAPAWSPTTYLLLLLLLSSGLSGTQDCSFQHSPISSDFAVK
                     IRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTVAGSKMQGLLERVN
                     TEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQP
                     DSSTLPPPWSPRPLEATAPTAPQPPLLLLLLLPVGLLLLAAAWCLHWQRTRRRTPRPG
                     EQVPPVPSPQDLLLVEH"
```

```
     sig_peptide     84..161
     misc_feature    162..629
                     /note="extracellular domain"
     misc_feature    630..698
                     /note="transmembrane domain"
     misc_feature    699..788
                     /note="cytoplasmic domain"
     3'UTR           792..1080
     misc_feature    1015..1080
                     /note="ATTTA mRNA instability motif"
     polyA_signal    1059..1064
     polyA_site      1080
                     /note="32 A residues"
BASE COUNT      204 a    384 c    290 g    202 t
ORIGIN

Initial Score    =    972  Optimized Score  =    979  Significance = 67.45
Residue Identity =    98%  Matches          =    980  Mismatches   =     8
Gaps             =      2  Conservative Substitutions               =     0

             X         10        20        30        40        50        60
             CGGCCGGA--ATTCCGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCAA
             || | |||    |  ||||||||||||||||||||||||||||||||||||||||||||||
   CAAGACCCGGCGACAGGAGGCATGAGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCAA
         50  X     60        70        80        90       100       110

          70        80        90       100       110       120       130
   CAACCTATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAACACA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CAACCTATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAACACA
      120       130       140       150       160       170       180

        140       150       160       170       180       190       200
   GCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCA
    190       200       210       220       230       240       250

      210       220       230       240       250       260       270
   CCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAGCGCTGGA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAGCGCTGGA
  260       270       280       290       300       310       320       330

    280       290       300       310       320       330       340
   TGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACACT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACACT
          340       350       360       370       380       390       400

  350       360       370       380       390       400       410       420
   TTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TTGTCACCAAATGTGCCTTTCAGCCCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCC
        410       420       430       440       450       460       470

          430       440       450       460       470       480       490
   TGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGG
```



```
      480       490       500       510       520       530       540

         500       510       520       530       540       550       560
   AGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCACCCCCATGGAGTCCCCGGCCCCTGGAGGCCACAGCCC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   AGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCACCCCCATGGAGTCCCCGGCCCCTGGAGGCCACAGCCC
    550       560       570       580       590       600       610

      570       580       590       600       610       620       630
   CGACAGCCCCCGCAGCCCCCTCTGCTCCTCCTACTGCTGCTGCCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CGACAGCCCCCGCAGCCCCCTCTGCTCCTCCTACTGCTGCTGCCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCT
  620       630       640       650       660       670       680       690

    640       650       660       670       680       690       700
   GGTGCCTGCACTGGCAGAGGACGCGGCGGAGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GGTGCCTGCACTGGCAGAGGACGCGGCGGAGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTC
          700       710       720       730       740       750       760

  710       720       730       740       750       760       770       780
   CCCAGGACCTGCTGCTTGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCTTAAACAACGCAGTG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CCCAGGACCTGCTGCTTGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCTTAAACAACGCAGTG
        770       780       790       800       810       820       830

          790       800       810       820       830       840       850
   AGACAGACATCTATCATCCCATTTTACAGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   AGACAGACATCTATCATCCCATTTTACAGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGG
      840       850       860       870       880       890       900

        860       870       880       890       900       910       920
   ATGTATAGCCTGGACACAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCTTGGGCCCCTCTCATTCCCTCCCC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ATGTATAGCCTGGACACAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCTTGGGCCCCTCTCATTCCCTCCCC
    910       920       930       940       950       960       970

      930       940       950       960       970       980       X
   AGAATGGAGGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCAACTCTGAACAAAGCCCCCG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
   AGAATGGAGGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCAACTCTGAACAAAGCCCTTGCCCCCATG
  980       990      1000      1010      1020      1030      1040 X    1050

   AA
```

2. US-08-162-407-5 (1-988)
   HSU04806 Human FLT3/FLK2 ligand mRNA, complete cds.

```
LOCUS       HSU04806      859 bp    mRNA            PRI       11-MAY-1994
DEFINITION  Human FLT3/FLK2 ligand mRNA, complete cds.
ACCESSION   U04806
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 798)
```

```
  AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
            Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
            Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
            Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
            Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
  TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
            haematopoietic stem cells and is encoded by variant RNAs
  JOURNAL   Nature 368, 643-648 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 859)
  AUTHORS   Culpepper,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
            Institute, 901 California Avenue, Palo Alto, CA 94304, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 483844
FEATURES             Location/Qualifiers
     source          1..859
                     /clone="human Flt3/Flk2 ligand S86"
                     /clone_lib="SV48 cDNA"
                     /organism="Homo sapiens"
                     /cell_line="SV48 thymic stroma"
     CDS             93..800
                     /note="NCBI gi: 483845"
                     /codon_start=1
                     /product="FLT3/FLK2 ligand"
                     /translation="MTVLAPAWSPTTYLLLLLLLSSGLSGTQDCSFQHSPISSDFAVK
                     IRELSDYLLQDYPVTVASNLQDEELCGALWRLVLAQRWMERLKTVAGSKMQGLLERVN
                     TEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQP
                     DSSTLPPPWSPRPLEATAPTAPQPPLLLLLLLPVGLLLLAAAWCLHWQRTRRRTPRPG
                     EQVPPVPSPQDLLLVEH"
BASE COUNT      147 a    318 c    246 g    148 t
ORIGIN

Initial Score    =    747  Optimized Score  =    767  Significance = 51.32
Residue Identity =    96%  Matches          =    771  Mismatches   =    25
Gaps             =      5  Conservative Substitutions               =     0

            X         10        20        30        40        50        60
            CGGCCGGA--ATTCCGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCAA
            || | |||    |  ||||||||||||||||||||||||||||||||||||||||||||||
   CCTGACCCGGCGACAGGAGGCATGAGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCAA
          60 X      70        80        90       100       110       120

           70        80        90       100       110       120       130
   CAACCTATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAACACA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CAACCTATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAACACA
       130       140       150       160       170       180       190

            140       150       160       170       180       190       200
   GCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCA
     200       210       220       230       240       250       260

       210       220       230       240       250       260       270
   CCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAGCGCTGGA
   ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
   CCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCGCTCTGGCGGCTGGTCCTGGCACAGCGCTGGA
```



```
     270       280       290       300       310       320       330

      280       290       300       310       320       330       340
   TGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACACT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACACT
 340       350       360       370       380       390       400       410

       350       360       370       380       390       400       410       420
   TTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TTGTCACCAAATGTGCCTTTCAGCCCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCC
         420       430       440       450       460       470       480

         430       440       450       460       470       480       490
   TGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGG
       490       500       510       520       530       540       550

        500       510       520       530       540       550       560
   AGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCACCCCCATGGAGTCCCCGGCCCCTGGAGGCCACAGCCC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   AGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCACCCCCATGGAGTCCCCGGCCCCTGGAGGCCACAGCCC
     560       570       580       590       600       610       620

      570       580       590       600       610       620       630
   CGACAGCCCCGCAGCCCCCTCTGCTCCTCCTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CGACAGCCCCGCAGCCCCCTCTGCTCCTCCTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCT
   630       640       650       660       670       680       690

    640       650       660       670       680       690       700
   GGTGCCTGCACTGGCAGAGGACGCGGCGGAGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GGTGCCTGCACTGGCAGAGGACGCGGCGGAGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTC
700       710       720       730       740       750       760       770

 710       720       730       740       750       760       770       780
   CCCAGGACCTGCTGCTTGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGCGGGAGCCTTAAACAACGCAGTG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CCCAGGACCTGCTGCTTGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGGGGAGGAT--ACGTAGGCACAC
         780       790       800       810       820       830       840

          790       800
   AGACAGACATCTATCATCCCATTTTACAG
   |||  |    || | || ||
   AGAGGGGAGTC-ACCAGCC
         850       X
```

3. US-08-162-407-5 (1-988)
   MMU04807 Mus musculus FLT3/FLK2 ligand mRNA, complete cds.

```
LOCUS       MMU04807     1152 bp    mRNA            ROD       11-MAY-1994
DEFINITION  Mus musculus FLT3/FLK2 ligand mRNA, complete cds.
ACCESSION   U04807
KEYWORDS    .
SOURCE      mouse.
  ORGANISM  Mus musculus
```

```
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Rodentia; Myomorpha; Muridae; Mus.
  REFERENCE 1  (bases 1 to 921)
    AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
            Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
            Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
            Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
            Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
    TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
            haematopoietic stem cells and is encoded by variant RNAs
    JOURNAL   Nature 368, 643-648 (1994)
    STANDARD  full automatic
  REFERENCE 2  (bases 1 to 1152)
    AUTHORS   Culpepper,J.A.
    TITLE     Direct Submission
    JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
            Institute, 901 California Avenue, Palo Alto, CA 94304, USA
    STANDARD  full automatic
  COMMENT     NCBI gi: 483846
  FEATURES             Location/Qualifiers
       source          1..1152
                       /clone="mouse Flt3/Flk2 ligand T110"
                       /clone_lib="TA4 cDNA"
                       /organism="Mus musculus"
                       /cell_line="TA4 stroma"
       CDS             256..954
                       /note="NCBI gi: 483847"
                       /codon_start=1
                       /product="FLT3/FLK2 ligand"
                       /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                       KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                       NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLGLKPCIGKACQNFSRCLEVQ
                       CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLLPLTLVLLAAAWGLRWQRARR
                       RGELHPGVPLPSHP"
  BASE COUNT      238 a    346 c    324 g    244 t
  ORIGIN

Initial Score    =    400  Optimized Score  =    611  Significance = 26.45
Residue Identity =    67%  Matches          =    646  Mismatches   =   258
Gaps             =     57  Conservative Substitutions               =     0

            X         10        20        30        40        50
            CGG---CCGGAATTCCGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCA
               |    | || | |  |||| |||||||| || ||||||||||||||||||||||||||||
   TTCCAAGAACCTGTCACAGGCA-TGAGGGGTCCCCGGCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCA
      220    X 230         240       250       260       270       280

   60          70        80        90        100       110       120
    ACAACCTATC---TCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAA
    |   |||  |   | || ||| ||||||||||||||  |  | ||  | |||||  |  |||||  | ||||
    AATTCCTCCCTGTTGCTGCTGTTGCTGCTGCTGAGTCCTTGCCTGCGGGGGGACACCTGACTGTTACTTCAGC
      290       300       310       320       330       340       350

   130       140       150       160       170       180       190       200
    CACAGCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCA
    ||||| ||||||||||| |||||   || ||  |  | ||| || ||||| ||||||| ||||||||||
    CACAGTCCCATCTCCTCCAACTTCAAAGTGAAGTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCA
    360       370       380       390       400       410       420

          210       220       230       240       250       260       270
```



```
   GTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAGCGC
   ||||| ||||||  ||| || |||||||| ||| ||||  || | | ||| | ||  |||| || ||||||
   GTCACTGTGGCCGTCAATCTTCAGGACGAGAAGCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGC
430        440       450       460       470       480       490       500

       280       290       300       310       320       330       340
   TGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGC-AAGGCTTGCTGGAGCGCGTGAACACGGAGAT
   ||||| ||||  || |||||||| || ||||| |||||||  || ||||  ||||||  ||| ||||| |||||
   TGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATGCAAACGCTT-CTGGAGGACGTCAACACCGAGAT
         510       520       530       540       550       560       570

      350       360       370       380       390       400       410
   ACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCG
   |||  |||||||||  ||||  |||| |||||||  ||     ||||| || ||||||||||||||||||||
   ACATTTTGTCACCTCATGTACCTTCCAGCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCA
        580       590       600       610       620       630       640

     420       430       440       450       460               470       480
   CCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATC-----ACTC--GCCAGAACTTCTC
   ||||||| |||| |||| |    |||||| | |  |||||||||||| |||     |  |  |||||||  |||||
   CCTCCTGAAGGACACCTGCACACAGCTGCTTGGTCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTC
      650       660       670       680       690       700       710

         490       500       510       520       530       540       550
   CCGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCACCCCCATGGAGTCCCCGGCCCCTGGA
    ||||||||||||| |||||||  ||||| ||||||||  |||||||  ||||||  ||||||||     |||| ||
   TCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGA
    720       730       740       750       760       770       780

         560       570           580       590       600           610
   GGCCAC--AGCCCCGACAGCCCCG---CAGCCCCCT-CTGCTCCTCCTACTGCTGCTGC--CC-GTGGGCCT
    |||||  |||  || |  |||| ||    |||  |  ||  |||||||  ||  ||||||||||  ||  |    ||
   AGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCTGCCTCTCACACT
790        800       810       820       830       840       850       860

    620       630       640       650       660       670       680
   CCTGCTGCTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCGGAGGACACCCCGCC-CTGGGGAGC
     ||||||||||  ||  |||||  |||| |  |||||  |||  |   |  |||||        | ||  |   ||
   GGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGG---
          870       880       890       900       910       920

   690       700       710       720       730       740       750
   AGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGC
    ||||||||  |  ||  |||| ||||    |   ||||| | |  ||||   |||| || || |||  |
   -GGTGCCCCTCCCCTCCCATCCCTAGGATGCGAGCCTTGTGCATCGTTGACTCAGCCAGGGTCTTATCTCGA
930        940       950       960       970       980       990      1000

760       770       780       790       800       810       820       830
   GGAGCCTTAAACAACGCAGTGAGACAGACATCTATCATCCCATTTTACAGGGGAGGATACTGAGGCACACAG
    |||   | ||          | || ||| |  |  | | | |  ||  ||||      | |   | |||  |||| |
   TGAGGTCTCAATATGTTGCCCAAACTGACTTTGAAAACCTCGAT-------GCACCTTCCTGCCCCACAAAC
         1010      1020      1030      1040              1050      1060

       840       850       860       870       880       890       900
   AGGGGAGTCACCAGCCAGAGGATGTATAGCCTGGACACAGAGGAAGTTGGCTAGAGGCCGGTC-CCTTCCTT
         |  || || | | || | ||  | |  | | | |  ||  ||||      |  ||  ||| |||
   TTCCAA---AC-AG-CTG-GGCT-TACGGGCATGCTATATA-CAACAAGGCT---TTCTTTTCTTCTTTCTT
    1070              1080      1090      1100              1110      1120

        910       920       930       940
```

```
    GGGCCCCTCTCATTCCCTCCCCAGAATGGAGGCAACGCCAGAATC
    ||    ||    ||         |||   |  |||
    GG--TGCTAGAGTT-------GGGAACCAAAACAA
      1130             1140      1150 X
```

4. US-08-162-407-5 (1-988)
   MUSLIGAND Mouse flt3 ligand mRNA, complete cds.

```
LOCUS       MUSLIGAND     829 bp ss-mRNA            ROD       04-JAN-1994
DEFINITION  Mouse flt3 ligand mRNA, complete cds.
ACCESSION   L23636
KEYWORDS    ligand.
SOURCE      Mus musculus (strain SJL/J) cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
REFERENCE   1  (bases 1 to 829)
  AUTHORS   Lyman,S.D., James,L., VandenBos,T., deVries,P., Brasel,K.,
            Gliniak,B., Hollingsworth,L., Picha,K.S., McKenna,H.J.,
            Splett,R.R., Fletcher,F.A., Maraskovsky,E., Farrah,T.,
            Foxworthe,D., Williams,D.E. and Beckmann,M.P.
  TITLE     Molecular cloning of a ligand for the flt3/flk-2 tyrosine kinase
            receptor -- a proliferative factor for primitive hematopoietic
            cells
  JOURNAL   Cell 75, 1157-1167 (1993)
  STANDARD  full automatic
COMMENT     NCBI gi: 439441
FEATURES             Location/Qualifiers
     source          1..829
                     /organism="Mus musculus"
                     /strain="SJL/J"
                     /sequenced_mol="cDNA to mRNA"
     sig_peptide     32..109
                     /codon_start=1
     CDS             32..727
                     /note="NCBI gi: 439442"
                     /codon_start=1
                     /product="flt3 ligand"
                     /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                     KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                     NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLALKPCIGKACQNFSRCLEVQ
                     CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLPLTLVLLAAAWGLRWQRARRR
                     GELHPGVPLPSHP"
     mat_peptide     110..724
                     /codon_start=1
BASE COUNT      163 a    265 c    221 g    180 t
ORIGIN

Initial Score    =    378  Optimized Score  =    580  Significance = 24.87
Residue Identity =    71%  Matches          =    612  Mismatches   =   193
Gaps             =     55  Conservative Substitutions               =     0

    X          10        20        30        40        50        60
    CGG---CCGGAATTCCGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCAACAACCTATC
    | |    | || | |  |||| ||||||| || ||||||||||||||||||||||||||||||||   |||  |
    CTGTCACAGGCA-TGAGGGGTCCCCGGCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCC
    X       10          20        30        40        50        60        70

      70        80        90       100       110       120       130
```



```
   ---TCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAACACAGCCCCA
      | || ||| ||||||||||||||  |  | ||  | |||||  |  |||||  | ||||   ||||| ||||
   TGTTGCTGCTGTTGCTGCTGCTGAGTCCTTGCCTGCGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCA
           80        90       100       110       120       130       140

  140       150       160       170       180       190       200       210
   TCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCACCGTGG
   |||||||| |||||   || ||  |  | ||| || ||||| |||||||| ||||||||||||||||| ||||
   TCTCCTCCAACTTCAAAGTGAAGTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGG
       150       160       170       180       190       200       210

           220       230       240       250       260       270       280
   CCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGC
   ||  ||| || |||||||||| ||| ||||  || | | ||| | ||  |||| || |||||||||| ||||
   CCGTCAATCTTCAGGACGAGAAGCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGC
     220       230       240       250       260       270       280

         290       300       310       320       330       340       350
   GGCTCAAGACTGTCGCTGGGTCCAAGATGC-AAGGCTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTC
     || |||||||| || ||||| ||||||| || |||| ||||||  ||| ||||| |||||||| ||||||
   AACTGAAGACTGTGGCAGGGTCTAAGATGCAAACGCTT-CTGGAGGACGTCAACACCGAGATACATTTTGTC
   290       300       310       320       330       340       350

       360       370       380       390       400       410       420
   ACCAAATGTGCCTTTCAGCCCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAG
   |||  |||| |||| |||||||  ||    ||||| || ||||||||||||||||||||||| ||||||| ||
   ACCTCATGTACCTTCCAGCCCCTACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAG
  360       370       380       390       400       410       420       430

    430       440       450       460               470       480       490
   GAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATC-----ACTC--GCCAGAACTTCTCCCGGTGCCTG
   || |||| |   |||||| | || |||||||||| |||    |  |  |||||| ||||| |||||||||
   GACACCTGCACACAGCTGCTTGCTCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTG
           440       450       460       470       480       490       500

         500       510       520       530       540       550       560
   GAGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCACCCCCATGGAGTCCCCGGCCCCTGGAGGCCAC--AG
   ||| ||||||| |||||  |||||||| |||||||  ||||||  ||||||||     |||| || |||||  ||
   GAGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAG
        510       520       530       540       550       560       570

         570           580       590       600       610       620
   CCCCGACAGCCCCG---CAGCCCCCT-CTGCTCCTCCTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCC
   | || | |||| ||    ||| |  ||  |||||||| || ||||||||||| |     ||  ||||||||||
   CTCCCAGAGCCTCGGCCCAGGCAGCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACTGGTGCTGCTGGCA
      580       590       600       610       620       630       640

  630       640       650       660       670       680       690
   GCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCGGA--GGACACC-CCGCCCTGGGGAGCAGGTGCCCCCC
   || |||||| |||| | |||||| ||| |  |  |||  ||  | | || |||||||| ||     || |||
   GCCGCCTGGGGCCTTCGCTGGCAAAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGGGGTGC----CCCTCCC
    650       660       670       680       690       700           710

  700       710       720       730       740       750       760       770
   GTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCTTAAA
    ||||   |||| |||    |  |||||| | |  ||||   |||| || ||  |||   |||    ||| |
   CTCCC--ATCCCTAGGATTCGAGCCTTGTGCATCGTTGACTCAGCCAGGGTCTTATC--TCGG----TTACA
        720       730       740       750       760           770

          780       790       800       810       820       830     X 840
```

```
CAACGCAGTGA-GACAGACATCTATCATCCCATTTTACAGGGGAGGA-TACTGAGGCACACAGAGGGGAGTC
C----CTGTAATCTCAGCCCT-TGGGAGCC--------CAGAGCAGGATTGCTGA------ATGGTCTGG
   780       790       800              810       820             X

ACCAGC
```

5. US-08-162-407-5 (1-988)
S65267 drebrin A {alternatively spliced, clone Dcb21} [ch

LOCUS S65267 1919 bp mRNA VRT 01-NOV-1993
DEFINITION drebrin A {alternatively spliced, clone Dcb21} [chickens, postnatal 1-day old, mRNA Partial, 1919 nt].
ACCESSION S65267
KEYWORDS .
SOURCE chickens postnatal 1-day old.
ORGANISM Gallus sp.
Unclassified.
REFERENCE 1 (bases 1 to 1919)
AUTHORS Kojima,N., Shirao,T. and Obata,K.
TITLE Molecular cloning of a developmentally regulated brain protein, chicken drebrin A and its expression by alternative splicing of the drebrin gene
JOURNAL Brain Res. Mol. Brain Res. 19, 101-114 (1993)
STANDARD full automatic
COMMENT GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 137017] from the original journal article.

This sequence comes from Fig. 3b.

NCBI gi: 410593
FEATURES Location/Qualifiers
source 1..1919
/organism="Gallus sp."
/note="chickens"
CDS 1..1565
/partial
/note="Authors give incomplete codon for residue number 20; Method: conceptual translation supplied by author. This sequence comes from Fig. 3b. Author-given protein sequence is in conflict with the conceptual translation.
NCBI gi: 410594"
/codon_start=1
/product="drebrin A"
/translation="LSNGLARVSSPVLHRLRLREDENAEPVGTTYQKTDATVEMKRLN
REQFWEQAKKEEELRKEEERKKALDARLRFEQERMEQERLEQEERERRYREREEQIEE
HRRKQQSMEAEEARQRLKEQSIFGEQQEEDDRQQLRKSESEVEEAAAIIAQRPDNPRE
FFKQQERVASGSGDAISPGSHRTGRLHCPFIKTADSGPPSSSSSSSSPPRTPFPYITC
HRTPNLSSFFPCSQSDYRKVSAAGCSPCESSPASTPLGEQRTRAPAEETPATPKDSPS
PSTQVAEPAATEQHWPFPGPEDKAAEPPGDEPDPDPRPAWTAGADVLGDLVTLEPSEP
SPAPAASEPQPVETPGVAEPLIELWQSDGAAPAATSTWPLPDTPAGPPVPPEEGTLLG
LDELPEPPATFCDAEQHEEVEEEEEEEEEATAGEPHPTGLGYQEGYQEGPEVPPITNG
EMGPKDGTAGRGEQASEGYFSQSQEEEAPPPEEPSAKAPQPVFYNKPPEIDITCWDTD
PLPEEEESFGGGL"
BASE COUNT 358 a 678 c 644 g 239 t
ORIGIN

Initial Score = 221 Optimized Score = 404 Significance = 13.62

Residue Identity = 49% Matches = 471 Mismatches = 376
Gaps = 100 Conservative Substitutions = 0

```
     90       100       110       120       130       140       150
TGAGCTCGGGACT-CAGTGGGACCCAGGACTGCTCCTTCCAACAC--AGCCCCATCTCCTCCGACTTC--GC
            TCTCCAACGGG---CTGGCGCGCGCGTC-TCCAGCCCCGTGCTGCACCGCCTGCGGCTGCGTGA
            X       10        20        30        40        50

    160       170       180       190       200       210       220
TGTC-AAAAT-CCGTGAGCTGTCTG-ACTACCTGCTTCAAGATTACCCA-GTCACCGTGGCCTCCAACCTGC
GGACGAGAATGCCGAG-CCCGTGGGCACGACCTAC--CA-GAAAACCGACGCCACCGTGGAGATGAAGCGGC
60        70        80        90       100       110       120

    230       240       250       260       270       280       290
AGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTG
TCAACCGGGAGCAGTTCTGGGAACAAGCCAAGAAAGAGGAGGAGTTGCGCAAGGAGGAGGAGGAGGAA-AAAG
130       140       150       160       170       180       190

   300       310       320       330       340       350       360
TCGCTGGGT-CCAAGATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCC
GCGTTGGATGCGCGGCTGCGGTTCGAGCAGGAGCGCATGGAGCAGGAGCGGC----TGGAGCAGGAGGAGCG
200       210       220       230       240       250       260

  370       380       390       400       410       420       430
TTTCAGCCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAG
CGAGAGGCGCTACC-----GGGAGCGCGAGGAGCAGATCGAGGAGCACAGGAGGAAGCAGCAGAGCATGGAG
 270       280       290       300       310       320       330

440       450       460       470       480       490       500
CAGCTGGTGGCGCTGAAGC-CCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAG--TGTCAG
GCGGAGGAGGCCCGGCAGCGCCTGAAGGA---GC-AGTCCATCTTTGGGGAGCAGCAAGAGGAGGACGACAG
     340       350       360       370       380       390       400

   510       520       530       540       550       560       570
CC--CGACTCCTCAACCCTGCCACCCCCAT---GGAGTCCCCGGCCCCTGGAGGCCACAGCCCCGACAGCCC
GCAGCAGCTCCGGAAATCAGAGTCAGAGGTGGAGGAGGCCGCTGCCATCATTGCCCAGCGACCTGACAACCC
       410       420       430       440       450       460       470

       580       590       600       610       620       630
C--GCAGCCCCCTCTGCTCCTCCTACTGCTG-----CTGCCCGTGG-GCCTCCTGCT-GC-TGGCCGC-----T
CCGGGAGTTCTTCAAGCAGCAGGAGCGGGTGGCATCGGGCAGCGGCGACGCCATCTCGCCGGGCAGCCATAG
     480       490       500       510       520       530       540

       640       650       660       670       680       690
GCCTGGT-GCCTGCACTGGC---AGAGGACG-CGGCGGAGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCG
GACAGGTCGTCTGCACTGTCCTTTCATAAAGACAGCTGACAGTGGGCCGCCTTCTTCTTC--TTCCTCCTCC
   550       560       570       580       590       600       610

700       710       720       730       740       750       760
TCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCA-CTGACCTGGCCAAGGCCTCATC-CTGCGGAGCCTTAA
```

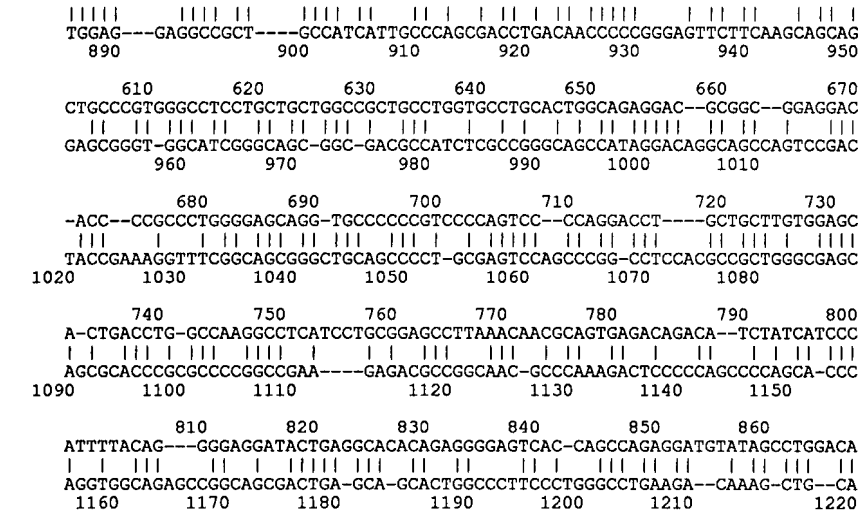
```
TCCTCCCCCCCGCGGACCCCTTTCCCCTATATCACCTGCCACCGC--ACCCCAAATCTCTCCTCTTTCTT--
   620       630       640       650       660       670       680

770       780       790       800       810       820       830
ACAACGCAGTGAGACAGACATCTATC-ATCCCATTTTACAG-GGGAGG-ATA--CTGAGGCACACAGAGGGG
CCCATGCAGCCAGTCCGA---CTACCGAAAGGTTTCGGCAGCGGGCTGCAGCCCCTGCGAGTC-CAGCCCGG
     690       700       710       720       730       740       750

  840       850       860       870       880       890       900
AGT-CAC-CAGCCAGAGGA--TGTATAGCCTGGACACAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCTTGG
CCTCCACGCCGCTGGGCGAGCAGCGCACCCGCGCCCCGGCCGAAG------AGACGCCGGCAACGCCCAAAG
      760       770       780       790               800       810

   910       920             930       940       950       960
GCCCCTCTCATTCCC--TCCCCAG-----AATGGAGGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCA
ACTCC-CCCAGCCCCAGCACCCAGGTGGCAGAGCCGGCAGCGACTGA--GCAGCACTGGCCC---TTCCCTG
 820       830       840       850       860       870       880

970       980       X
ACTCTGA--ACAAAGCCCCCG
   ||||  |||||||  | |
GGCCTGAAGACAAAGCTGCAGAGCCGCCGGG
     890       900  X    910
```

6. US-08-162-407-5 (1-988)
   CHKDEBPDA Chicken drebrin mRNA.

```
LOCUS       CHKDEBPDA    2102 bp ss-mRNA           VRT       27-JUL-1990
DEFINITION  Chicken drebrin mRNA.
ACCESSION   M36961
KEYWORDS    drebrin.
SOURCE      Chicken brain cDNA to mRNA, clone pDcw17.
  ORGANISM  Gallus gallus
            Eukaryota; Animalia; Chordata; Vertebrata; Aves; Neornithes;
            Neognathae; Galliformes; Phasianidae.
REFERENCE   1  (bases 1 to 2102)
  AUTHORS   Kojima,N., Kato,Y., Shirao,T. and Obata,K.
  TITLE     Nucleotide sequences of two embryonic drebrins, developmentally
            regulated brain proteins, and developmental change in their mRNAs
  JOURNAL   Mol. Brain Res. 4, 207-215 (1988)
  STANDARD  full automatic
COMMENT     NCBI gi: 211725
FEATURES             Location/Qualifiers
     source          1..2102
                     /organism="Gallus gallus"
     CDS             60..1883
                     /note="drebrin;  NCBI gi: 211726"
                     /codon_start=1
                     /translation="MAGVGFAAHRLELLASYQDVIGEDSPTDWALYTYEDGSDDLKLA
                     ASGGGGLLELSGHFEIQKVMYGFCSVKEPQAVLPKYVLVNWVGEDVPDARKCACASHV
                     AKIAEFFQGVDVIVNASSVEDIDPGAIGQRLSNGLARVSSPVLHRLRLREDENAEPVG
                     TTYQKTDATVEMKRLNREQFWEQAKKEEELRKEEERKKALDARLRFEQERMEQERLEQ
                     EERERRYREREEQIEEHRRKQQSMEAEEARQRLKEQSIFGEQQEEDDRQQLRKSESEV
                     EEAAAIIAQRPDNPREFFKQQERVASGSGDAISPGSHRTGSQSDYRKVSAAGCSPCES
                     SPASTPLGEQRTRAPAEETPATPKDSPSPSTQVAEPAATEQHWPFPGPEDKAAEPPGD
                     EPDPDPRPAWTAGADVLGDLVTLEPSEPSPAPAASEPQPVETPGVAEPLIELWQSDGA
```



```
                    APAATSTWPLPDTPAGPPVPPEEGTLLGLDELPEPPATFCDAEQHEEVEEEEEEEEEA
                    TAGEPHPTGLGYQEGYQEGPEVPPITNGEMGPKDGTAGRGEQASEGYFSQSQEEEAPP
                    PEEPSAKAPQPVFYNKPPEIDITCWDTDPLPEEEESFGGGL"
BASE COUNT      399 a    707 c    742 g    254 t
ORIGIN

Initial Score    =    221  Optimized Score  =    448  Significance = 13.62
Residue Identity =    50%  Matches          =    536  Mismatches   =   400
Gaps             =    126  Conservative Substitutions               =     0

            X          10          20          30          40          50
            CGGCC---GGAATTC-CGGGGCC--CCCGGCCGA--AATGACAGTG-CTGG-CGCCAGCCTG
GCCCGCAAATGTGCCTGTGCCAGCCACGTGGCCAAGATCGCAGAGTTCTTCCAGGGTGTGGATGTTATCGTC
  340     X 350       360       370       380       390       400

      60          70        80          90         100         110
GAGCCCAACAACCT--ATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGAC-CCA--GGACT--
AATGCCAGCAGCGTGGAGGACAT--TGACCCGGGGGC---CATCGGG---CAGCGGCTCTCCAACGGGCTGG
410       420       430         440               450       460       470

 120         130         140         150         160         170         180
-GCTCCT-TCCAACAC--AGCCCCATCTCCTCCGACTTC--GCTGTC-AAAAT-CCGTGAGCTGTCTG-ACT
CGCGCGTCTCCAGCCCCGTGCTGCACCGCCTGCGGCTGCGTGAGGACGAGAATGCCGAG-CCCGTGGGCACG
      480       490         500         510         520         530       540

       190         200         210         220         230         240         250
ACCTGCTTCAAGATTACCCA-GTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTCTGG
ACCTAC--CA-GAAAACCGACGCCACCGTGGAGATGAAGCGGCTCAACCGGGAGCAGTTCTGGGAACAAGCC
       550         560         570         580         590         600       610

      260         270         280         290         300         310         320
CGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGT-CCAAGATGCAAGGCTTGCT
AAGAAAGAGGAGGAGTTGCGCAAGGAGGAGGAGAGGAA-AAAGGCGTTGGATGCGCGGCTGCGGTTCGAGCA
      620         630         640         650         660         670       680

     330         340         350         360         370         380         390
GGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTT
GGAGCGCATGGAGCAGGAGCGGC----TGGAGCAGGAGGAGCGCGAGAGGCGCTACC-----GGGAGCGCGA
     690         700         710         720         730         740

   400         410         420         430         440         450         460
CGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGC-CCTGGATCA
GGAGCAGATCGAGGAGCACAGGAGGAAGCAGCAGAGCATGGAGGCGGAGGAGGCCCGGCAGCGCCTGAAGGA
  750         760         770         780         790         800         810

  470         480         490         500             510         520         530
CTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAG--TGTCAGCC--CGACTCCTCAACCCTGCCACCCCCA
---GC-AGTCCATCTTTGGGGAGCAGCAAGAGGAGGACGACAGGCAGCAGCTCCGGAAATCAGAGTCAGAGG
    820         830         840         850         860         870         880

    540         550         560         570         580         590         600
TGGAGTCCCCGGCCCCTGGAGGCCA-CAGCCCCGACAGCCCCG-CAGCCCCC-TCTGCTCCTCCTACTGCTG
```

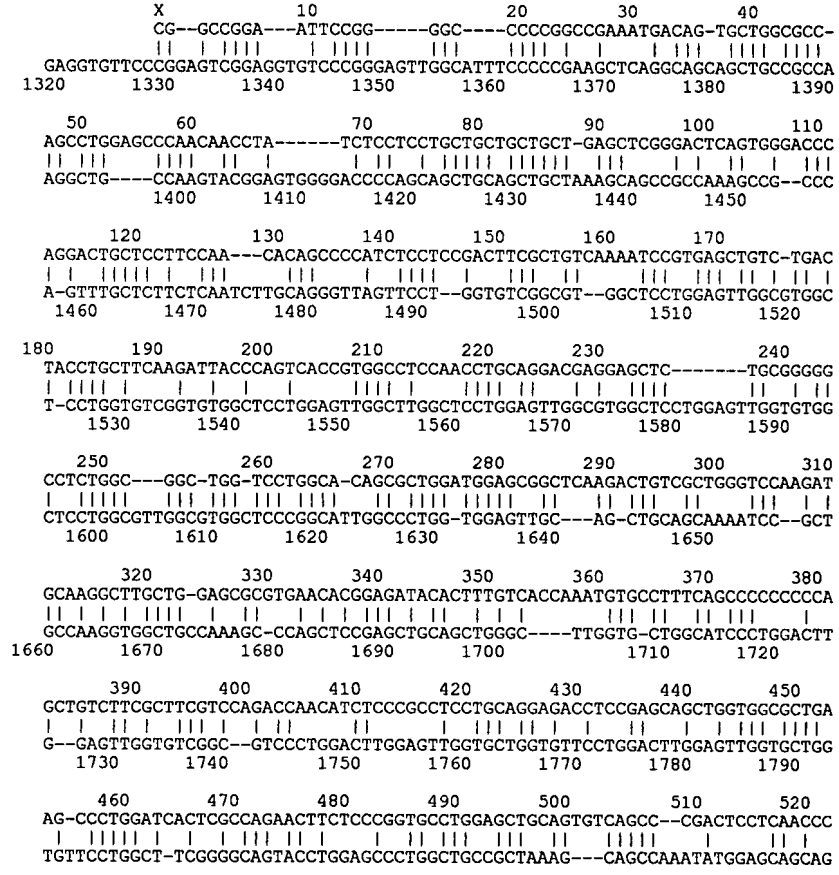


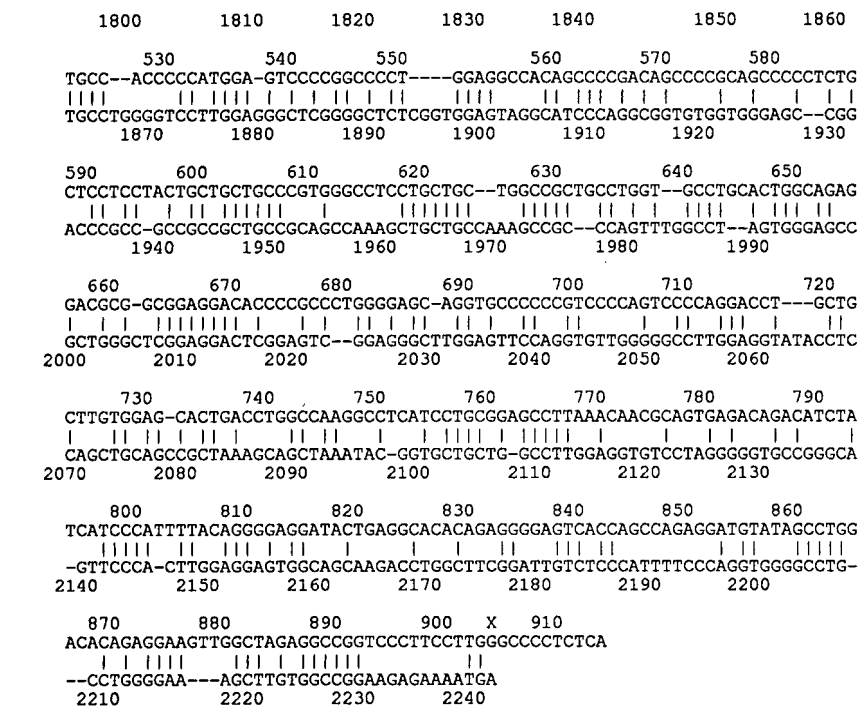
7. US-08-162-407-5 (1-988)
   HSELASRNA H.sapiens RNA for dermal fibroblast elastin.

LOCUS HSELASRNA 2241 bp RNA PRI 16-DEC-1994
DEFINITION H.sapiens RNA for dermal fibroblast elastin.
ACCESSION X52896
KEYWORDS elastin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE 1 (bases 1 to 2241)
AUTHORS Fazio,M.J., Olsen,D.R., Kauh,E.A., Baldwin,C.T., Indik,Z., Ornstein-Goldstein,N., Yeh,H., Rosenbloom,J. and Uitto,J.
TITLE Cloning of full-length elastin cDNAs from a human skin fibroblast recombinant cDNA library: further elucidation of alternative

splicing utilizing exon-specific oligonucleotides
JOURNAL J. Invest. Dermatol. 91, 458-464 (1988)
STANDARD full automatic
COMMENT NCBI gi: 31133
FEATURES Location/Qualifiers
source 1..2241
/organism="Homo sapiens"
BASE COUNT 333 a 597 c 840 g 471 t
ORIGIN

Initial Score = 218 Optimized Score = 405 Significance = 13.40
Residue Identity = 49% Matches = 478 Mismatches = 374
Gaps = 113 Conservative Substitutions = 0



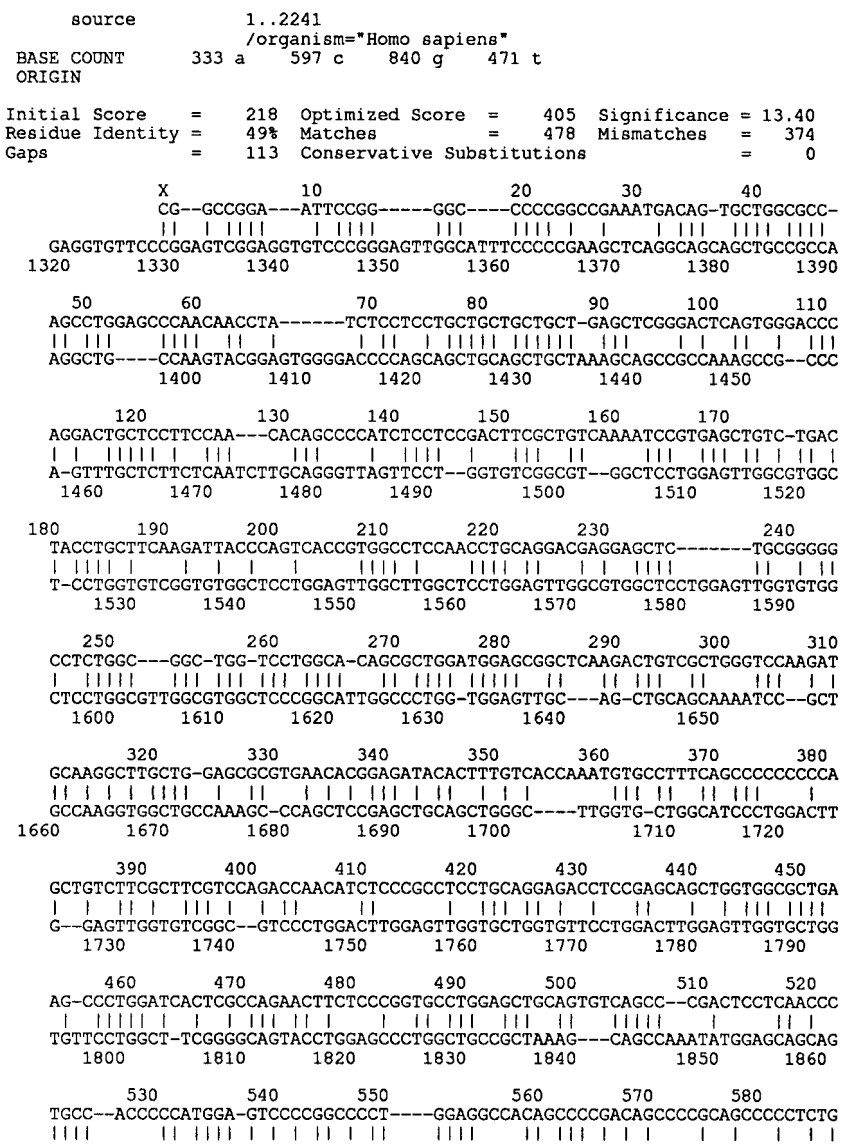


8. US-08-162-407-5 (1-988)
HSELASRNA H.sapiens RNA for dermal fibroblast elastin.

LOCUS HSELASRNA 2241 bp RNA PRI 23-NOV-1994
DEFINITION H.sapiens RNA for dermal fibroblast elastin.
ACCESSION X52896
KEYWORDS elastin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE 1 (bases 1 to 2241)
AUTHORS Fazio,M.J., Olsen,D.R., Kauh,E.A., Baldwin,C.T., Indik,Z., Ornstein-Goldstein,N., Yeh,H., Rosenbloom,J. and Uitto,J.
TITLE Cloning of full-length elastin cDNAs from a human skin fibroblast recombinant cDNA library: further elucidation of alternative splicing utilizing exon-specific oligonucleotides
JOURNAL J. Invest. Dermatol. 91, 458-464 (1988)
STANDARD full automatic
COMMENT NCBI gi: 31133
FEATURES Location/Qualifiers

source 1..2241
/organism="Homo sapiens"
BASE COUNT 333 a 597 c 840 g 471 t
ORIGIN

Initial Score = 218 Optimized Score = 405 Significance = 13.40
Residue Identity = 49% Matches = 478 Mismatches = 374
Gaps = 113 Conservative Substitutions = 0



```
TGCCTGGGGTCCTTGGAGGGCTCGGGGCTCTCGGTGGAGTAGGCATCCCAGGCGGTGTGGTGGGAGC--CGG
     1870      1880      1890      1900      1910      1920      1930

590       600       610       620       630       640       650
CTCCTCCTACTGCTGCTGCCCGTGGGCCTCCTGCTGC--TGGCCGCTGCCTGGT--GCCTGCACTGGCAGAG
  || ||  | || ||||||   |      |||||||   |||||  || | |  ||||  | ||| ||
ACCCGCC-GCCGCCGCTGCCGCAGCCAAAGCTGCTGCCAAAGCCGC--CCAGTTTGGCCT--AGTGGGAGCC
      1940      1950      1960      1970      1980      1990

  660       670       680       690       700       710       720
GACGCG-GCGGAGGACACCCCGCCCTGGGGAGC-AGGTGCCCCCCGTCCCCAGTCCCCAGGACCT---GCTG
|  | |  |||||||| |   | |  || | ||  || |  ||  ||      |  ||  |||  |    ||
GCTGGGCTCGGAGGACTCGGAGTC--GGAGGGCTTGGAGTTCCAGGTGTTGGGGGGCCTTGGAGGTATACCTC
2000      2010      2020      2030      2040      2050      2060

      730       740       750       760       770       780       790
CTTGTGGAG-CACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCTTAAACAACGCAGTGAGACAGACATCTA
|   || || | || |    || ||   |   | |||| | |||||  |      |   | |   | |    |
CAGCTGCAGCCGCTAAAGCAGCTAAATAC-GGTGCTGCTG-GCCTTGGAGGTGTCCTAGGGGGTGCCGGGCA
2070      2080      2090      2100      2110      2120      2130

     800       810       820       830       840       850       860
TCATCCCATTTTACAGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGGATGTATAGCCTGG
   ||||||  ||  ||| | ||   |      |   |   ||   ||| ||          | ||   |||||
-GTTCCCA-CTTGGAGGAGTGGCAGCAAGACCTGGCTTCGGATTGTCTCCCATTTTCCCAGGTGGGGCCTG-
2140      2150      2160      2170      2180      2190      2200

  870       880       890       900   X   910
ACACAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCTTGGGCCCCTCTCA
   | | ||||    ||| | ||||||         ||
--CCTGGGGAA---AGCTTGTGGCCGGAAGAGAAAATGA
 2210      2220      2230      2240
```

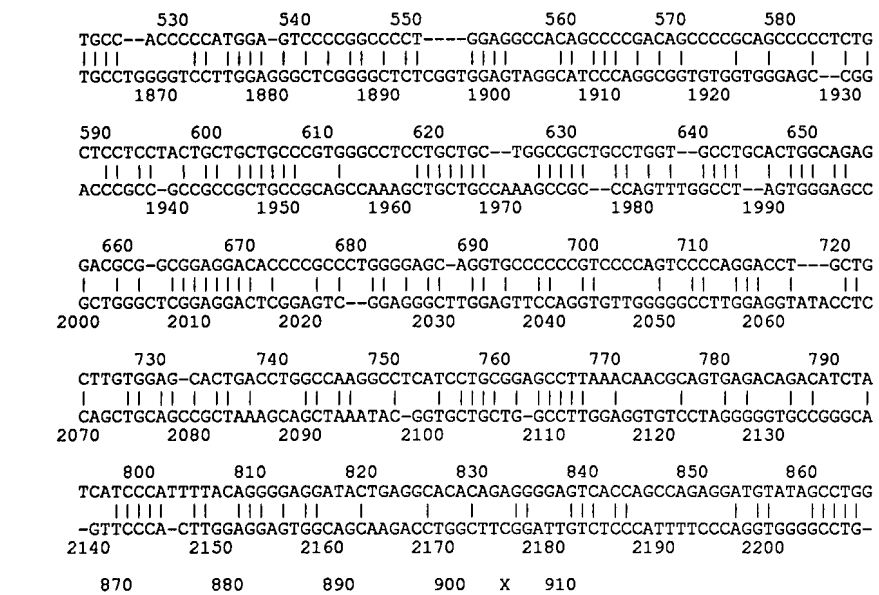
9. US-08-162-407-5 (1-988)
HSELASRNA H.sapiens RNA for dermal fibroblast elastin

```
ID   HSELASRNA  standard; RNA; PRI; 2241 BP.
XX
AC   X52896;
XX
DT   18-FEB-1992 (Rel. 31, Created)
DT   16-DEC-1994 (Rel. 42, Last updated, Version 5)
XX
DE   H.sapiens RNA for dermal fibroblast elastin
XX
KW   elastin.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
XX
RN   [1]
RP   1-2241
RA   Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
RA   Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.;
RT   "Cloning of full-length elastin cDNAs from a human skin fibroblast
RT   recombinant cDNA library: further elucidation of alternative
RT   splicing utilizing exon-specific oligonucleotides";
RL   J. Invest. Dermatol. 91:458-464(1988).
```



```
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2241
FT                   /organism="Homo sapiens"
XX
SQ   Sequence 2241 BP; 333 A; 597 C; 840 G; 471 T; 0 other;
```

```
Initial Score    =    218  Optimized Score  =    405  Significance = 13.40
Residue Identity =    49%  Matches          =    478  Mismatches   =   374
Gaps             =    113  Conservative Substitutions              =     0

            X           10                  20        30        40
            CG--GCCGGA---ATTCCGG-----GGC----CCCCGGCCGAAATGACAG-TGCTGGCGCC-
              ||  | ||||    | ||||     |||    |||| |  |    | |||  |||| ||||
  GAGGTGTTCCCGGAGTCGGAGGTGTCCCGGGAGTTGGCATTTCCCCCGAAGCTCAGGCAGCAGCTGCCGCCA
1320      1330      1340      1350      1360      1370      1380      1390

   50        60              70        80        90       100       110
  AGCCTGGAGCCCAACAACCTA------TCTCCTCCTGCTGCTGCTGCT-GAGCTCGGGACTCAGTGGGACCC
  || |||    ||||  ||  |       | ||  | ||||| ||||||  |||    | |  ||  |  |||
  AGGCTG-----CCAAGTACGGAGTGGGGACCCCAGCAGCTGCAGCTGCTAAAGCAGCCGCCAAAGCCG--CCC
           1400      1410      1420      1430      1440      1450

        120         130       140       150       160       170
  AGGACTGCTCCTTCCAA---CACAGCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTC-TGAC
  | |  ||||| |  |||     |||     |  ||||  |   |||  ||      |||  ||| || | || |
  A-GTTTGCTCTTCTCAATCTTGCAGGGTTAGTTCCT--GGTGTCGGCGT--GGCTCCTGGAGTTGGCGTGGC
   1460      1470      1480      1490      1500      1510      1520

 180       190       200       210       220       230             240
  TACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTC-------TGCGGGGG
  | |||| |    |  |  |   |    ||||  |   |||| ||   | |  ||||         ||  | ||
  T-CCTGGTGTCGGTGTGGCTCCTGGAGTTGGCTTGGCTCCTGGAGTTGGCGTGGCTCCTGGAGTTGGTGTGG
     1530      1540      1550      1560      1570      1580      1590

    250             260       270       280       290       300       310
  CCTCTGGC---GGC-TGG-TCCTGGCA-CAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGAT
  |  |||||   ||| ||| ||| ||||   || |||| |||||  ||   || |||  ||    |||  | |
  CTCCTGGCGTTGGCGTGGCTCCCGGCATTGGCCCTGG-TGGAGTTGC---AG-CTGCAGCAAAATCC--GCT
    1600      1610      1620      1630      1640           1650

        320       330       340       350       360       370       380
  GCAAGGCTTGCTG-GAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCA
  || | | | ||||  |  ||   | | | ||| | ||  | | |        ||| ||  || |||    |
  GCCAAGGTGGCTGCCAAAGC-CCAGCTCCGAGCTGCAGCTGGGC----TTGGTG-CTGGCATCCCTGGACTT
1660      1670      1680      1690      1700           1710      1720

      390       400       410       420       430       440       450
  GCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGA
  |  |  || |  ||| |  | ||      ||       |  ||| || |  |   ||   |  | ||| ||||
  G--GAGTTGGTGTCGGC--GTCCCTGGACTTGGAGTTGGTGCTGGTGTTCCTGGACTTGGAGTTGGTGCTGG
     1730      1740      1750      1760      1770      1780      1790

     460       470       480       490       500       510       520
  AG-CCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCC--CGACTCCTCAACCC
   |  ||||| |  |  | ||| || |    |  || |||  |||  ||   |||||    |    || |
  TGTTCCTGGCT-TCGGGGCAGTACCTGGAGCCCTGGCTGCCGCTAAAG---CAGCCAAATATGGAGCAGCAG
     1800      1810      1820      1830      1840           1850      1860
```

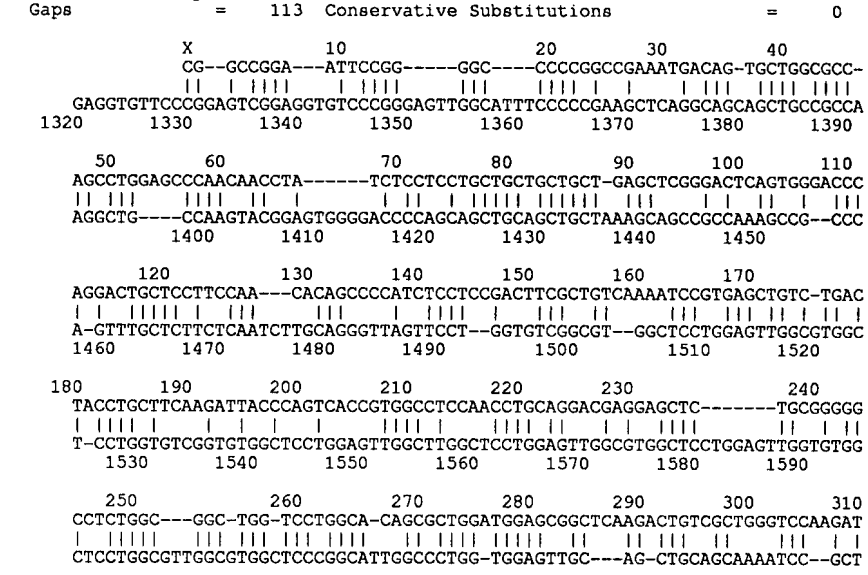


10. US-08-162-407-5 (1-988)
HUMELASF Human elastin mRNA, complete cds.

LOCUS HUMELASF 2242 bp ss-mRNA PRI 07-NOV-1994
DEFINITION Human elastin mRNA, complete cds.
ACCESSION M36860
KEYWORDS elastin.
SOURCE Human skin fibroblast, cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE 1 (bases 1 to 2242)
AUTHORS Fazio,M.J., Olsen,D.R., Kauh,E.A., Baldwin,C.T., Indik,Z., Ornstein-Goldstein,N., Yeh,H., Rosenbloom,J. and Uitto,J.
TITLE Cloning of full-length elastin cDNAs from a human skin fibroblast recombinant cDNA library: further elucidation of alternative splicing utilizing exon-specific oligonucleotides
JOURNAL J. Invest. Dermatol. 91, 458-464 (1988)
STANDARD full automatic
COMMENT NCBI gi: 182061
FEATURES Location/Qualifiers
source 1..2242
/organism="Homo sapiens"

sig_peptide 50..127
/note="elastin signal peptide"
/codon_start=1
CDS 50..2242
/gene="ELN"
/map="7cen-q21.1"
/note="elastin precursor; NCBI gi: 182062"
/codon_start=1
/gdb_xref="GO0-119-107"
/translation="MAGLTAAAPRPGVLLLLLSILHPSRPGGVPGAIPGGVPGGVFYP
GAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVTFPGALVPGGVADAAAA
YKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPGVYPGGVLPGARF
PGVGVLPGVPTGAGVKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGYPIKAPKLPGGYG
LPYTTGKLPYGYGPGGVAGAAGKAGYPTGTGVGPQAAAAAAAKAAAKFGAGAAGVLPG
VGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVV
GVPGAGVPGVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVG
GIPTYGVGAGGFPGFGVGVGGIPGVAGVPSVGGVPGVGGVPGVGISPEAQAAAAAKAA
KYGVGTPAAAAAKAAAKAAQFALLNLAGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVG
VAPGVGVAPGVGVAPGIGPGGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVP
GLGVGAGVPGLGVGAGVPGFGAVPGALAAAAKAAKYGAAVPGVLGGLGALGGVGIPGGV
VGAGPAAAAAAAKAAAKAAQFGLVGAAGLGGLGVGGLGVPGVGGLGGIPPAAAAKAAK
YGAAGLGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGKACGRKRK"
mat_peptide 128..2239
/note="elastin"
/codon_start=1
BASE COUNT 333 a 598 c 840 g 471 t
ORIGIN

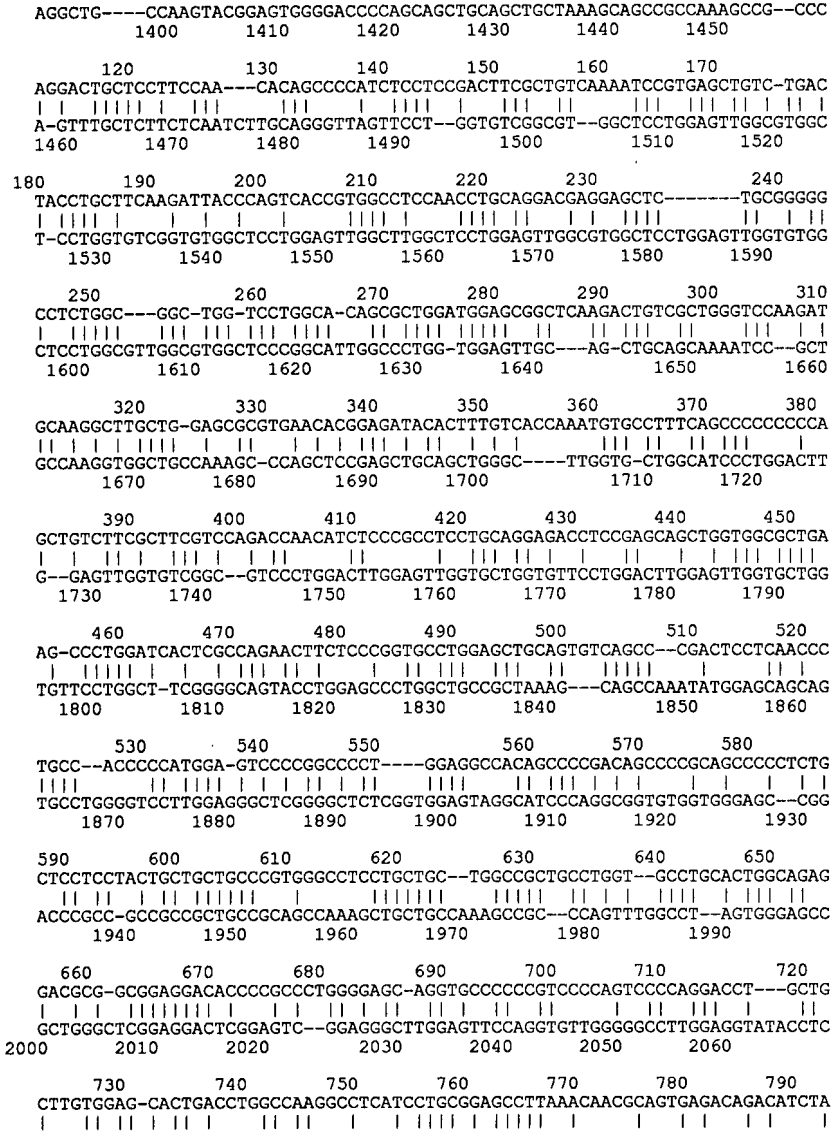
Initial Score = 218 Optimized Score = 405 Significance = 13.40
Residue Identity = 49% Matches = 478 Mismatches = 374
Gaps = 113 Conservative Substitutions = 0

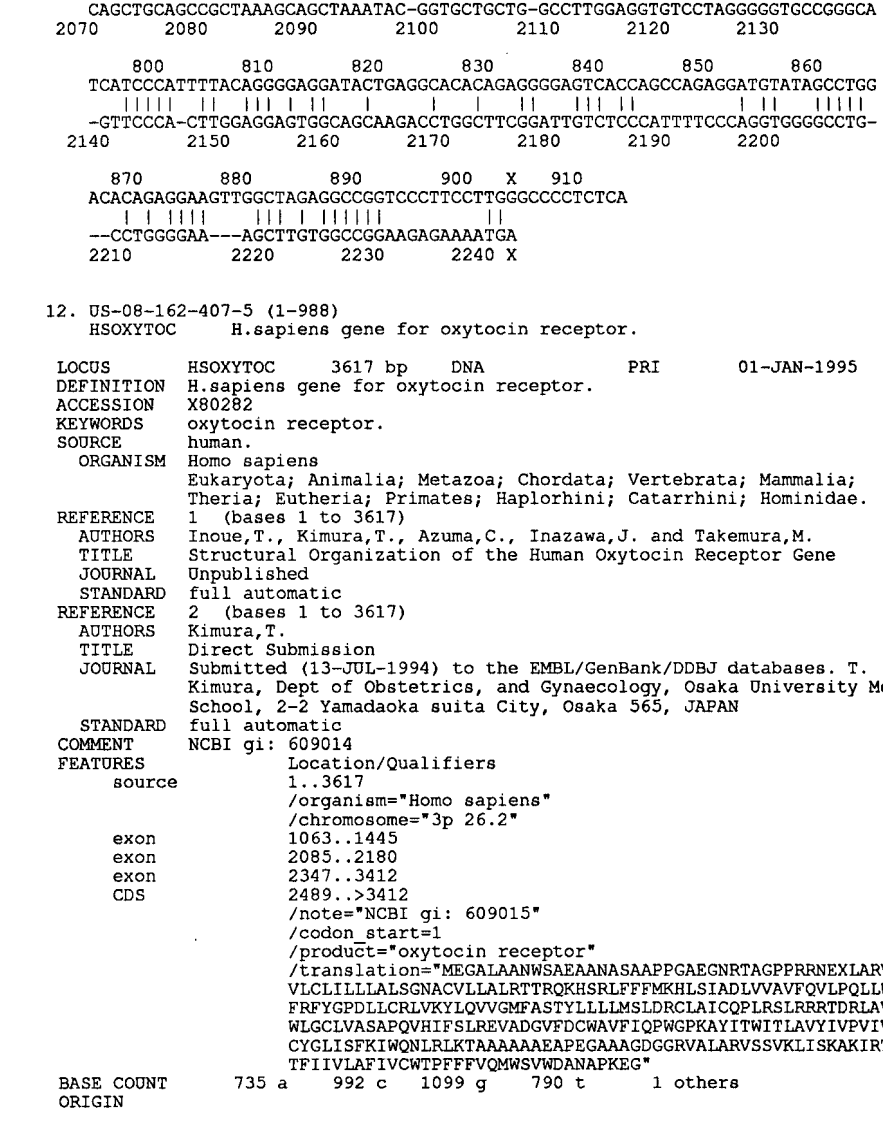


```
 1600       1610       1620       1630       1640           1650        1660

         320        330        340        350        360        370        380
GCAAGGCTTGCTG-GAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCA
|| | | | |||| |  ||   | | | ||| | ||  | | |       ||| ||  || |||    |
GCCAAGGTGGCTGCCAAAGC-CCAGCTCCGAGCTGCAGCTGGGC----TTGGTG-CTGGCATCCCTGGACTT
       1670       1680       1690       1700           1710       1720

       390        400        410        420        430        440        450
GCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGA
|  |  || |  ||| |  | ||     ||       |  ||| || |  |   ||   |  | ||| ||||
G--GAGTTGGTGTCGGC--GTCCCTGGACTTGGAGTTGGTGCTGGTGTTCCTGGACTTGGAGTTGGTGCTGG
  1730       1740       1750       1760       1770       1780       1790

     460        470        480        490        500        510        520
AG-CCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCC--CGACTCCTCAACCC
  |  ||||| |   |  | ||| || |     |  || |||  |||  ||   |||||     |      || |
TGTTCCTGGCT-TCGGGGCAGTACCTGGAGCCCTGGCTGCCGCTAAAG---CAGCCAAATATGGAGCAGCAG
  1800       1810       1820       1830       1840           1850       1860

       530        540        550            560        570        580
TGCC--ACCCCCATGGA-GTCCCCGGCCCCT----GGAGGCCACAGCCCCGACAGCCCCGCAGCCCCCTCTG
||||      || |||| | | | || | ||     ||||     || ||| | | |     |  |    |  | |
TGCCTGGGGTCCTTGGAGGGCTCGGGGCTCTCGGTGGAGTAGGCATCCCAGGCGGTGTGGTGGGAGC--CGG
    1870       1880       1890       1900       1910       1920       1930

590        600        610        620        630        640        650
CTCCTCCTACTGCTGCTGCCCGTGGGCCTCCTGCTGC--TGGCCGCTGCCTGGT--GCCTGCACTGGCAGAG
  || ||   | ||  ||||||    |       |||||||     |||||  || | |  ||||  | ||| ||
ACCCGCC-GCCGCCGCTGCCGCAGCCAAAGCTGCTGCCAAAGCCGC--CCAGTTTGGCCT--AGTGGGAGCC
     1940       1950       1960       1970       1980       1990

  660        670        680        690        700        710          720
GACGCG-GCGGAGGACACCCCGCCCTGGGGAGC-AGGTGCCCCCCGTCCCCAGTCCCCAGGACCT---GCTG
|  | |  |||||||| |   | |  || | ||  || |  ||  ||     |  ||  |||  |    ||
GCTGGGCTCGGAGGACTCGGAGTC--GGAGGGCTTGGAGTTCCAGGTGTTGGGGGCCTTGGAGGTATACCTC
2000       2010       2020       2030       2040       2050       2060

       730        740        750        760        770        780        790
CTTGTGGAG-CACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCTTAAACAACGCAGTGAGACAGACATCTA
|   || || | || |    || ||   |    | |||| | ||||  |      |   | |   | |    |
CAGCTGCAGCCGCTAAAGCAGCTAAATAC-GGTGCTGCTG-GCCTTGGAGGTGTCCTAGGGGGTGCCGGGCA
2070       2080       2090       2100       2110       2120       2130

     800        810        820        830        840        850        860
TCATCCCATTTTACAGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGGATGTATAGCCTGG
   |||||  ||  ||| | ||   |    |   |   ||   ||| ||          | ||   |||||
-GTTCCCA-CTTGGAGGAGTGGCAGCAAGACCTGGCTTCGGATTGTCTCCCATTTTCCCAGGTGGGGCCTG-
2140       2150       2160       2170       2180       2190       2200

  870        880        890        900   X   910
ACACAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCTTGGGCCCCTCTCA
   | | ||||    ||| | ||||||          ||
--CCTGGGGAA---AGCTTGTGGCCGGAAGAGAAAATGA
2210       2220       2230       2240 X


11. US-08-162-407-5 (1-988)
    HSELASF      Human elastin mRNA, complete cds.
```



```
ID   HSELASF    standard; RNA; PRI; 2242 BP.
XX
AC   M36860;
XX
DT   05-SEP-1990 (Rel. 25, Created)
DT   16-DEC-1994 (Rel. 42, Last updated, Version 5)
XX
DE   Human elastin mRNA, complete cds.
XX
KW   elastin.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
XX
RN   [1]
RP   1-2242
RA   Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
RA   Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.;
RT   "Cloning of full-length elastin cDNAs from a human skin fibroblast
RT   recombinant cDNA library: further elucidation of alternative
RT   splicing utilizing exon-specific oligonucleotides";
RL   J. Invest. Dermatol. 91:458-464(1988).
XX
DR   SWISS-PROT; P15502; ELS_HUMAN.
XX
CC   NCBI gi: 182061
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2242
FT                   /organism="Homo sapiens"
FT   sig_peptide     50..127
FT                   /note="elastin signal peptide"
FT                   /codon_start=1
FT   CDS             50..2242
FT                   /gene="ELN"
FT                   /map="7cen-q21.1"
FT                   /note="elastin precursor; NCBI gi: 182062"
FT                   /codon_start=1
FT                   /gdb_xref="G00-119-107"
FT   mat_peptide     128..2239
FT                   /note="elastin"
FT                   /codon_start=1
XX
SQ   Sequence 2242 BP; 333 A; 598 C; 840 G; 471 T; 0 other;

Initial Score    =    218  Optimized Score  =    405  Significance = 13.40
Residue Identity =    49%  Matches          =    478  Mismatches   =   374
Gaps             =    113  Conservative Substitutions               =     0

            X           10                  20         30         40
            CG--GCCGGA---ATTCCGG-----GGC----CCCCGGCCGAAATGACAG-TGCTGGCGCC-
            ||  | ||||    | ||||     |||    |||| |  |    | |||  |||| ||||
   GAGGTGTTCCCGGAGTCGGAGGTGTCCCGGGAGTTGGCATTTCCCCCGAAGCTCAGGCAGCAGCTGCCGCCA
1320      1330       1340       1350       1360       1370       1380      1390

     50         60              70         80         90         100        110
   AGCCTGGAGCCCAACAACCTA------TCTCCTCCTGCTGCTGCTGCT-GAGCTCGGGACTCAGTGGGACCC
   || |||    ||||  ||  |       | ||  | ||||| ||||||  |||    | |  ||  |  |||
```





12. US-08-162-407-5 (1-988)
HSOXYTOC H.sapiens gene for oxytocin receptor.

LOCUS HSOXYTOC 3617 bp DNA PRI 01-JAN-1995
DEFINITION H.sapiens gene for oxytocin receptor.
ACCESSION X80282
KEYWORDS oxytocin receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE 1 (bases 1 to 3617)
AUTHORS Inoue,T., Kimura,T., Azuma,C., Inazawa,J. and Takemura,M.
TITLE Structural Organization of the Human Oxytocin Receptor Gene
JOURNAL Unpublished
STANDARD full automatic
REFERENCE 2 (bases 1 to 3617)
AUTHORS Kimura,T.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-1994) to the EMBL/GenBank/DDBJ databases. T. Kimura, Dept of Obstetrics, and Gynaecology, Osaka University Med School, 2-2 Yamadaoka suita City, Osaka 565, JAPAN
STANDARD full automatic
COMMENT NCBI gi: 609014
FEATURES Location/Qualifiers
source 1..3617
/organism="Homo sapiens"
/chromosome="3p 26.2"
exon 1063..1445
exon 2085..2180
exon 2347..3412
CDS 2489..>3412
/note="NCBI gi: 609015"
/codon_start=1
/product="oxytocin receptor"
/translation="MEGALAANWSAEAANASAAPPGAEGNRTAGPPRRNEXLARVEVA
VLCLILLLALSGNACVLLALRTTRQKHSRLFFFMKHLSIADLVVAVFQVLPQLLWDIT
FRFYGPDLLCRLVKYLQVVGMFASTYLLLLMSLDRCLAICQPLRSLRRRTDRLAVLAT
WLGCLVASAPQVHIFSLREVADGVFDCWAVFIQPWGPKAYITWITLAVYIVPVIVLAA
CYGLISFKIWQNLRLKTAAAAAAEAPEGAAAGDGGRVALARVSSVKLISKAKIRTVKM
TFIIVLAFIVCWTPFFFVQMWSVWDANAPKEG"
BASE COUNT 735 a 992 c 1099 g 790 t 1 others
ORIGIN

```
Initial Score    =    193  Optimized Score  =    461  Significance = 11.61
Residue Identity =    52%  Matches          =    561  Mismatches   =   361
Gaps             =    141  Conservative Substitutions               =     0

             X       10        20        30        40        50        60
             CGGCCGGAATTCCGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAG--CCCAA
             |  ||| |   ||||  ||||  | || || ||  |  |||||||  || |||||    |
   GGGCCGAGGGCAACCGCACCGCCGGACCCCCGCGGCGCAACGA-GGSCCTGGCGC--GCGTGGAGGTGGCGG
      2560  X   2570      2580      2590      2600      2610      2620

           70        80        90       100       110       120
   CAACCTATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACC--CAGG-ACTGCTCCTTCCAAC
        | |||| |||||||| |||  ||||||  ||||| |  ||| |   |  | ||  |||| |    ||
   TGCTGTGTCTCATCCTGCTCCTGGCGCTGAG--CGGGAATGCGTGTGTGCTGCTGGCGCTGCGCACCACACG
        2630      2640      2650      2660      2670      2680      2690

  130       140       150       160       170               180
    ACAGCCCCATCTCCTCCGACTTCGCTGTCA-AAATC-CGTGAGC-TGTCTGAC-------TACC--TGCTTC
     |||   || |||   |  ||||   |||  || | | | ||| |  | |||       |  |  || |||
    CCAGAAGCA-CTCGCGCCTCTTCTTCTTCATGAAGCACCTAAGCATCGCCGACCTGGTGGTGGCAGTGTTTC
        2700      2710      2720      2730      2740      2750      2760

  190         200       210         220       230       240       250
    AAG-ATTACC-CAGTCACCGTGG--CCTCCAACCTGCAG-GACGAGGAGCTCTGCGGGGGCCTCTGGCGGCT
    | |   | || ||||  | ||||  | ||  ||||  | |  | | | || |  |  |   || || || ||
    AGGTGCTGCCGCAGTTGCTGTGGGACATC--ACCTTCCGCTTCTACGGGCCCGACCTG---CTGTGCCGCCT
        2770      2780      2790      2800      2810      2820

     260       270       280       290       300       310       320
    GGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTG---CTGGA
    ||| |  | ||   || || ||| | |||  | |   | | |   | |  | ||| |   | ||   |||||
    GGT-CAAGTAC-TTGCAGG-TGGTG-GGC--ATGTTCGCCTCCACCTACCTGCTGCTGCTCATGTCCCTGGA
  2830        2840      2850      2860      2870      2880      2890

       330       340       350       360       370       380       390
    GCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTTCGT
     ||| ||  |  ||  ||   ||  | || |   || | ||  | || || || || | | |  ||| | |
    CCGC-TG--CCTGGCCAT---CT--GCCAGCCGCTGCG-CTCGCTGCGCCGCCGCA-CCGAC--CGCCTGG-
      2900      2910          2920      2930      2940      2950

    400       410       420       430       440       450       460
    CCAGACCAACATCTCCCGCCTC--CTG-CAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATCAC
     |||  |  | | |  | |||  ||| | || | ||  || || || ||| || |   | |      || |
    -CAGTGC-TCGCCACGTGGCTCGGCTGCCTGGTGGCCAGCGCGCCGCAGGT-GCAC---ATC-----TTCTC
       2960      2970      2980      2990      3000          3010

      470       480       490       500       510       520       530
    TC-GCCAGAACT-TCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGAC-TCCTC-AACCCTGCCACCCCCAT
    || ||  ||  |  ||  ||| | ||    |||| |     | ||| | || || | |||||     ||||
    TCTGCGCGAGGTGGCTGACGGCGTCTTCGACTGCTG-----GGCCGTCTTCATCCAGCCCTG-GGGACCCAA
     3020      3030      3040      3050          3060      3070      3080

        540       550       560          570       580       590
    GGAGTCCCCGGCCCCTGGAGGCCAC-AGC-CCCGACA--GCCCCG--CAGCCCCCTCTGCT-CCTCCTACTG
    ||  |   |  | | ||||   | | |||   | |||  |  |||  ||' |  ||| ||| ||| |||| |
    GGCCT--ACATCACATGGATCACGCTAGCTGTCTACATCGTGCCGGTCATCGTGCTC-GCTGCCTGCTACGG
          3090      3100      3110      3120      3130      3140      3150

   600       610       620       630       640       650       660
    -CTGCTGCCCGTGGGCCTCCTGCTG--CTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCG---G
```



```
     ||  |   | |     ||  || |  || || ||| |  |    |||||  ||| | || || ||||   |
    CCTTATCAGCTTCAAGATCTGGCAGAACTTGCGGCT-CAAGACCGCTGCAGCGGCGGCGGCCGAGGCGCCAG
          3160      3170      3180      3190      3200      3210      3220

      670       680       690       700       710       720       730
    AGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCAC-TG
    ||| | |  ||  |||| ||   ||  || | |   | ||| | | |   | || ||  ||  ||| | |
    AGGGCGCGGCG-GCTGGCGATGGGGGGCGCGTGG--CCCTGGCGC--GTGTCAGCAGC--GTCAAGCTCATC
          3230      3240      3250      3260      3270      3280

      740       750       760       770       780       790
    ACC-TGGCCAAG-GCCTCA---TCCTGCGGAGCCTTAAACAACGCAGTGAGACAGACATC-TATCAT---CC
     ||  |||||||  || ||   ||  |  || | || | || ||   || | |  |||| | |  |   ||
    TCCAAGGCCAAGATCCGCACGGTCAAGATGA-CTTTCATCATCGTGCTG-GCC-TTCATCGTGTGCTGGACC
    3290      3300      3310      3320      3330          3340      3350

  800       810       820       830       840       850       860
    CATTT-TAC-AGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGGATGTATAGCCTGGAC--
    | ||| | |   | |  |||   |||   |   ||  | |   ||  |||  |||| |  ||||| || |
    CCTTTCTTCTTCGTGCAGATGTGGAGCGTCTGGGATGCCA---ACGCGCCCAAGGAAG-GTAGCCAGGGCTG
     3360      3370      3380      3390          3400      3410      3420

    870       880       890       900       910       920       930
    ACAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCTTGGGCCCCTCTCATTCCCTCCCCAGAATG--GAGGCAA
      |||   ||  ||   ||| | |||  ||        |||| | || || |  || |||||||   |||
    GGAGACCCAGGAGGAGGGAGCCTGGTGGCTGGGGGAAGCCCTTATC-TTGCTGCCTCAGAATGTCCAGGGGT
      3430      3440      3450      3460      3470      3480      3490

    940       950           960           970           980       X
    CGCCAGA-ATCCAG---CACCGGCCCCATTTA---CCCA---ACT--CTGAAC--AAAGCCCCCG
    |    ||  ||| |   |  ||    || |  ||||  | |  |||  |    | |||  |
    CTGTGGACTTCCTGGGGGATAAGCGGGTTTGAAATCCCACAGAGTCACTGCTCTGTCATCCCTTGACCAAGT
      3500      3510      3520      3530      3540      3550      3560

    GAC
```

13. US-08-162-407-5 (1-988)
    HSOXYTOC H.sapiens gene for oxytocin receptor

```
ID   HSOXYTOC   standard; DNA; PRI; 3617 BP.
XX
AC   X80282;
XX
DT   01-JAN-1995 (Rel. 42, Created)
DT   01-JAN-1995 (Rel. 42, Last updated, Version 2)
XX
DE   H.sapiens gene for oxytocin receptor
XX
KW   oxytocin receptor.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
XX
RN   [1]
RA   Inoue T., Kimura T., Azuma C., Inazawa J., Takemura M.;
RT   "Structural Organization of the Human Oxytocin Receptor Gene";
RL   Unpublished.
```

```
XX
RN   [2]
RP   1-3617
RA   Kimura T.;
RT   ;
RL   Submitted (13-JUL-1994) to the EMBL/GenBank/DDBJ databases.
RL   T. Kimura, Dept of Obstetrics, and Gynaecology, Osaka University
RL   Med School, 2-2 Yamadaoka suita City, Osaka 565, JAPAN
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..3617
FT                   /organism="Homo sapiens"
FT                   /chromosome="3p 26.2"
FT   CDS             2489..>3412
FT                   /partial
FT                   /product="oxytocin receptor"
FT   exon            1063..1445
FT   exon            2085..2180
FT   exon            2347..3412
XX
SQ   Sequence 3617 BP; 735 A; 992 C; 1099 G; 790 T; 1 other;
```

```
Initial Score    =    193  Optimized Score  =    461  Significance = 11.61
Residue Identity =    52%  Matches          =    561  Mismatches   =   361
Gaps             =    141  Conservative Substitutions             =     0
```

```
            X       10        20        30        40        50          60
            CGGCCGGAATTCCGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAG--CCCAA
              |  ||| |   ||||  ||||  | || || ||  |  |||||||  || |||||    |
    GGGCCGAGGGCAACCGCACCGCCGGACCCCCGCGGCGCAACGA-GGSCCTGGCGC--GCGTGGAGGTGGCGG
       2560  X   2570      2580      2590      2600      2610      2620

           70        80        90       100       110       120
    CAACCTATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACC--CAGG-ACTGCTCCTTCCAAC
         | |||| |||||||| |||  ||||||  |||||  |  ||| |  |  | ||  |||| |    ||
    TGCTGTGTCTCATCCTGCTCCTGGCGCTGAG--CGGGAATGCGTGTGTGCTGCTGGCGCTGCGCACCACACG
         2630      2640      2650      2660      2670      2680      2690

  130       140       150       160       170       180
    ACAGCCCCATCTCCTCCGACTTCGCTGTCA-AAATC-CGTGAGC-TGTCTGAC--------TACC--TGCTTC
      |||   || |||   |  ||||   |||  || | | | ||| |  | |||        |  |  || |||
    CCAGAAGCA-CTCGCGCCTCTTCTTCTTCATGAAGCACCTAAGCATCGCCGACCTGGTGGTGGCAGTGTTTC
         2700      2710      2720      2730      2740      2750      2760

  190       200       210       220       230       240       250
    AAG-ATTACC-CAGTCACCGTGG--CCTCCAACCTGCAG-GACGAGGAGCTCTGCGGGGGCCTCTGGCGGCT
    | |   | || ||||  | ||||  | ||  |||| | |   | | | || |  |  |   || || || ||
    AGGTGCTGCCGCAGTTGCTGTGGGACATC--ACCTTCCGCTTCTACGGGCCCGACCTG---CTGTGCCGCCT
        2770      2780      2790      2800      2810      2820

     260       270       280       290       300       310       320
    GGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTG---CTGGA
    ||| |  | ||   || || ||| | |||  | |   | | |    | |  | |||    | ||   |||||
    GGT-CAAGTAC-TTGCAGG-TGGTG-GGC--ATGTTCGCCTCCACCTACCTGCTGCTGCTCATGTCCCTGGA
  2830        2840      2850      2860      2870      2880      2890

      330       340       350       360       370       380       390
    GCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTTCGT
     ||| ||  |  ||  ||   ||  | || |   || | ||  | || || || || | | |  ||| | |
```



```
CCGC-TG--CCTGGCCAT---CT--GCCAGCCGCTGCG-CTCGCTGCGCCGCCGCA-CCGAC--CGCCTGG-
  2900      2910      2920      2930      2940      2950

400       410       420       430       440       450       460
CCAGACCAACATCTCCCGCCTC--CTG-CAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATCAC
 |||  |  |  | |  | ||| ||| | || | ||  || || || ||| || |   | |     || |
-CAGTGC-TCGCCACGTGGCTCGGCTGCCTGGTGGCCAGCGCGCCGCAGGT-GCAC---ATC-----TTCTC
   2960      2970      2980      2990      3000          3010

  470       480       490       500       510       520       530
TC-GCCAGAACT-TCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGAC-TCCTC-AACCCTGCCACCCCCAT
|| ||  ||  |  ||  ||| | ||    |||| |     | ||| | || || | |||||     ||||
TCTGCGCGAGGTGGCTGACGGCGTCTTCGACTGCTG-----GGCCGTCTTCATCCAGCCCTG-GGGACCCAA
 3020      3030      3040      3050          3060      3070      3080

   540       550       560          570       580       590
GGAGTCCCCGGCCCCTGGAGGCCAC-AGC-CCCGACA--GCCCCG--CAGCCCCCTCTGCT-CCTCCTACTG
||  |   | | | ||||   | | |||   | |||  |  |||  || |   ||| ||| ||| |||| |
GGCCT--ACATCACATGGATCACGCTAGCTGTCTACATCGTGCCGGTCATCGTGCTC-GCTGCCTGCTACGG
       3090      3100      3110      3120      3130      3140      3150

600       610       620       630       640       650       660
-CTGCTGCCCGTGGGCCTCCTGCTG--CTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCG---G
 ||  |   | |      ||  || |  || || ||| |  |    |||||  ||| | || || ||||    |
CCTTATCAGCTTCAAGATCTGGCAGAACTTGCGGCT-CAAGACCGCTGCAGCGGCGGCGGCCGAGGCGCCAG
      3160      3170      3180      3190      3200      3210      3220

  670       680       690       700       710       720       730
AGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCAC-TG
||| | |  ||  |||| ||   || || |   |  ||| | | |   |  || ||  ||  ||| | |
AGGGCGCGGCG-GCTGGCGATGGGGGGCGCGTGG--CCCTGGCGC--GTGTCAGCAGC--GTCAAGCTCATC
      3230      3240      3250      3260      3270      3280

  740       750       760       770       780       790
ACC-TGGCCAAG-GCCTCA---TCCTGCGGAGCCTTAAACAACGCAGTGAGACAGACATC-TATCAT---CC
 ||  |||||||  || ||   ||  | || | || | || ||   || | |   |||| | | |   ||
TCCAAGGCCAAGATCCGCACGGTCAAGATGA-CTTTCATCATCGTGCTG-GCC-TTCATCGTGTGCTGGACC
3290      3300      3310      3320      3330      3340      3350

800       810       820       830       840       850       860
CATTT-TAC-AGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGGATGTATAGCCTGGAC--
| ||| | |   | |  |||   |||   |  || | |  ||  |||  |||| |  ||||| || | |
CCTTTCTTCTTCGTGCAGATGTGGAGCGTCTGGGATGCCA---ACGCGCCCAAGGAAG-GTAGCCAGGGCTG
 3360      3370      3380      3390          3400      3410      3420

870       880       890       900       910       920       930
ACAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCTTGGGCCCCTCTCATTCCCTCCCCAGAATG--GAGGCAA
  |||   ||  ||  ||| | |||  ||         |||| | || || |  || |||||||   |||
GGAGACCCAGGAGGAGGGAGCCTGGTGGCTGGGGGAAGCCCTTATC-TTGCTGCCTCAGAATGTCCAGGGGT
   3430      3440      3450      3460      3470      3480      3490

940       950       960       970       980       X
CGCCAGA-ATCCAG---CACCGGCCCCATTTA---CCCA---ACT--CTGAAC--AAAGCCCCCG
|    ||  |||  |   |   ||   ||  |  ||||   | |  |||  |    | |||  |
CTGTGGACTTCCTGGGGGATAAGCGGGTTTGAAATCCCACAGAGTCACTGCTCTGTCATCCCTTGACCAAGT
  3500      3510      3520      3530      3540      3550      3560

GAC
```

```
14. US-08-162-407-5 (1-988)
    HSMRNAOXY    H.sapiens mRNA for oxytocin receptor.

 LOCUS       HSMRNAOXY    4103 bp    RNA             PRI       10-NOV-1993
 DEFINITION  H.sapiens mRNA for oxytocin receptor.
 ACCESSION   X64878
 KEYWORDS    hormone receptor; oxytocin receptor.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
             Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
 REFERENCE   1  (bases 1 to 4103)
   AUTHORS   Kimura,T.
   TITLE     Direct Submission
   JOURNAL   Submitted (28-FEB-1992) to the EMBL/GenBank/DDBJ databases. T.
             Kimura, Dept of Obstetrics and Gynecology, Osaka University Medical
             School, 2-2 Yamadaoka, Suita-shi, Osaka 565, JAPAN
   STANDARD  full automatic
 REFERENCE   2  (bases 1 to 4100)
   AUTHORS   Kimura,T., Tanizawa,O., Mori,K., Brownstein,M.J. and Okayama,H.
   TITLE     Structure and expression of a human oxytocin receptor [published
             erratum appears in Nature 1992 May 14;357(6374):176]
   JOURNAL   Nature 356, 526-529 (1992)
   STANDARD  full automatic
 REFERENCE   3  (bases 1 to 4103)
   AUTHORS   Kimura,T.
   TITLE     Direct Submission
   JOURNAL   Submitted (28-APR-1992) to the EMBL/GenBank/DDBJ databases. Kimura
             T., same adress
   STANDARD  full automatic
 COMMENT     NCBI gi: 34764
 FEATURES             Location/Qualifiers
      source          1..4103
                      /organism="Homo sapiens"
                      /tissue_type="uterus"
      CDS             368..1537
                      /note="NCBI gi: 34765"
                      /codon_start=1
                      /product="oxytocin receptor"
                      /translation="MEGALAANWSAEAANASAAPPGAEGNRTAGPPRRNEALARVEVA
                      VLCLIILLALSGNACVLLALRTTRQKHSRLFFFMKHLSIADLVVAVFQVLPQLLWDIT
                      FRFYGPDLLCRLVKYLQVVGMFASTYLLLLMSLDRCLAICQPLRSLRRRTDRLAVLAT
                      WLGCLVASAPQVHIFSLREVADGVFDCWAVFIQPWGPKAYITWITLAVYIVPVIVLAT
                      CYGLISFKIWQNLRLKTAAAAAAEAPEGAAAGDGGRVALARVSSVKLISKAKIRTVKM
                      TFIIVLAFIVCWTPFFFVQMWSVWDANAPKEASAFIIVMLLASLNSCCNPWIYMLFTG
                      HLFHELVQRFLCCSASYLKGRRLGETSASKKSNSSSFVLSHRSSSQRSCSQPSTA"
 BASE COUNT     1171 a    966 c   1035 g    931 t
 ORIGIN

Initial Score    =    193  Optimized Score  =    456  Significance = 11.61
Residue Identity =    52%  Matches          =    552  Mismatches   =   373
Gaps             =    136  Conservative Substitutions              =     0

            X       10        20        30        40        50          60
            CGGCCGGAATTCCGGGGCCCCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAG--CCCAA
            |  ||| |   ||||  ||||  | || || ||  |  |||||||  || |||||    |
   GGGCCGAGGGCAACCGCACCGCCGGACCCCCGCGGCGCAACGA-GGCCCTGGCGC--GCGTGGAGGTGGCGG
         440 X     450       460       470       480       490       500
```



```
          70        80        90       100       110       120       130
  CAACCTATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACTC-AGTGGGACCCAGG-ACTGCTCCTTCCAACA
     | |||| ||||||| |||  |||||||  || || |  ||| |   | ||  |||| |    ||
  TGCTGTGTCTCATCCTGCTCCTGGCGCTGAGC-GGGAACGCGTGTGTGCTGCTGGCGCTGCGCACCACACGC
          510       520       530       540       550       560       570

          140       150       160       170                 180       190
  CAGCCCCATCTCCTCCGACTTCGCTGTCA-AAATC-CGTGAGC-TGTCTGAC-------TACC--TGCTTCA
  |||   || |||   |  ||||   |||  || | | | ||| |  | |||       |  |  || ||||
  CAGAAGCA-CTCGCGCCTCTTCTTCTTCATGAAGCACCTAAGCATCGCCGACCTGGTGGTGGCAGTGTTTCA
          580       590       600       610       620       630       640

            200       210       220       230       240       250
  AG-ATTACC-CAGTCACCGTGG--CCTCCAACCTGCAG-GACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTG
   |   | || |||| | ||||  | ||  |||| | |   | | | || |  |  |   || || || |||
  GGTGCTGCCGCAGTTGCTGTGGGACATC--ACCTTCCGCTTCTACGGGCCCGACCTG---CTGTGCCGCCTG
        650       660       670       680       690       700

  260       270       280       290       300       310       320
  GTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTG---CTGGAG
  || |  | ||   || || ||| | |||  | |   | | |     | |  | |||     | ||   |||||
  GT-CAAGTAC-TTGCAGG-TGGTG-GGC--ATGTTCGCCTCCACCTACCTGCTGCTGCTCATGTCCCTGGAC
710         720       730       740       750       760       770

   330       340       350       360       370       380       390
  CGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTTCGTC
  ||| ||  |  ||  ||   ||  | || |   || | ||  | || || || || | | |  ||| | |
  CGC-TG--CCTGGCCAT---CT--GCCAGCCGCTGCG-CTCGCTGCGCCGCCGCA-CCGAC--CGCCTGG--
     780             790       800       810       820       830

 400       410       420       430       440       450       460
  CAGACCAACATCTCCCGCCTC--CTG-CAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTGGATCACT
  |||  |  |  | |  | |||  ||| | || | ||  || || || ||| || |   | |       || ||
  CAGTGC-TCGCCACGTGGCTCGGCTGCCTGGTGGCCAGCGCGCCGCAGGT-GCAC---ATC-----TTCTCT
       840       850       860       870       880                 890

  470       480       490       500       510       520       530
  C-GCCAGAACT-TCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGAC-TCCTC-AACCCTGCCACCCCCATG
  | ||  ||  |  ||  ||| | ||    |||| |     | ||| | || || | |||||     |||| |
  CTGCGCGAGGTGGCTGACGGCGTCTTCGACTGCTG-----GGCCGTCTTCATCCAGCCCTG-GGGACCCAAG
    900       910       920       930               940       950       960

    540       550       560               570       580       590
  GAGTCCCCGGCCCCTGGAGGCCAC-AGC-CCCGACA--GCCCCG--CAGCCCCCTCTGCT-CCTCCTACTG-
  |  |   | | |||||   | | |||   | |||  |  |||  ||  |   ||| ||| ||| |||| |
  GCCT--ACATCACATGGATCACGCTAGCTGTCTACATCGTGCCGGTCATCGTGCTC-GCTACCTGCTACGGC
          970       980       990      1000      1010      1020      1030

 600       610       620       630       640       650       660
  CTGCTGCCCGTGGGCCTCCTGCTG--CTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCG---GA
  ||  |   | |     ||  || | || || ||| | |    |||||  ||| | || || ||||   ||
  CTTATCAGCTTCAAGATCTGGCAGAACTTGCGGCT-CAAGACCGCTGCAGCGGCGGCGGCCGAGGCGCCAGA
       1040      1050      1060      1070      1080      1090      1100

  670               680       690       700       710       720       730
  GGACAC-----CCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAG-GACCTGCTGCTTGTGGA--
  || | |     |  ||  ||||| ||  |||  ||| | | |  |||  ||| | |  ||| | | | |
  GGGCGCGGCGGCTGGCGATGGGGGGCGCGTG-GCCCTGGCGCGTGTCAGCAGCGTCAAGCT-CATCTCCAAG
       1110      1120      1130      1140      1150      1160      1170
```

```
            740         750         760         770         780         790
GCACTGA----CCTGGCCAAG--GCC-----TCATCCTGCGGAGCCTTAAAC-AACGCAGTGAGACAGACAT
||   ||     |  || ||||  | |     ||||| ||| | |||||  | |    || |   | |   | |
GCCAAGATCCGCACGGTCAAGATGACTTTCATCATCGTGCTG-GCCTTCATCGTGTGCTGGACGCCTTTCTT
     1180        1190        1200        1210        1220        1230        1240

        800         810         820         830         840         850
CTATCATCCCATTTTACAGGG--GAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGGAT-GT-AT
|| || | |    | |  || |    |||| |  | || | || || | || | |    |   |  || || ||
CT-TCGTGCAGATGTGGAGCGTCTGGGATGCCAACGCGCCCA-AG-GAAGCCTCGGCCTTCATCATCGTCAT
     1250        1260        1270        1280        1290        1300        1310

860         870         880         890         900         910         920
-AGCCTGGACA-CAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCT-TGGGCCCCTCTCATTCCCTCCCCAGA
   ||||| || |      || ||      |  || |   || | |   |  | |   |   | || ||  ||
GCTCCTGGCCAGCCTCAACAGCTGCTGCAACCCCTGGATCTACATGCTGTTCACGGGCCACCTCTTCCACGA
     1320        1330        1340        1350        1360        1370        1380

930         940         950         960         970         980         X
ATGGAGGCAACGCCAGAATCC--AGCACCGGCCCCATTTACCCAACTCTGAACAAAGCCCCCG
|     ||| |||     |||   ||  |  | |||  ||||  |    |  || |  ||  |
ACTCGTGCAGCGC----TTCCTGTGCTGCTCCGCCAGCTACCTGA---AGGGCAGACGCCTGGGAGAGACGA
   1390          1400        1410        1420          1430        1440

G
```

15. US-08-162-407-5 (1-988)
    TCU17917 Thermomonospora curvata alpha-glucosidase (aglA) g

```
LOCUS       TCU17917     2093 bp ds-DNA             BCT       19-DEC-1994
DEFINITION  Thermomonospora curvata alpha-glucosidase (aglA) gene, complete
            cds.
ACCESSION   U17917
KEYWORDS    .
SOURCE      Thermomonospora curvata.
  ORGANISM  Thermomonospora curvata
            Eubacteria; Firmicutes; Actinomycetes; Thermomonospora group;
            Thermomonospora.
REFERENCE   1  (bases 1 to 2093)
  AUTHORS   Janda,L., Tichy,P. and Petricek,M.
  TITLE     Thermomonospora curvata aglA gene encoding alpha-glucosidase
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 2093)
  AUTHORS   Petricek,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1994) Miroslav Petricek, Institute of
            Microbiology, ASCR, Biogenesis & Biotechnol. of Natur. Prod.,
            Videnska 1083, Prague 4, 142 20, Czech Republic
  STANDARD  full automatic
COMMENT     NCBI gi: 603047
FEATURES             Location/Qualifiers
     source          1..2093
                     /strain="CCM3352"
                     /organism="Thermomonospora curvata"
     CDS             109..1743
                     /gene="aglA"
                     /EC_number="3.2.1.20"
```

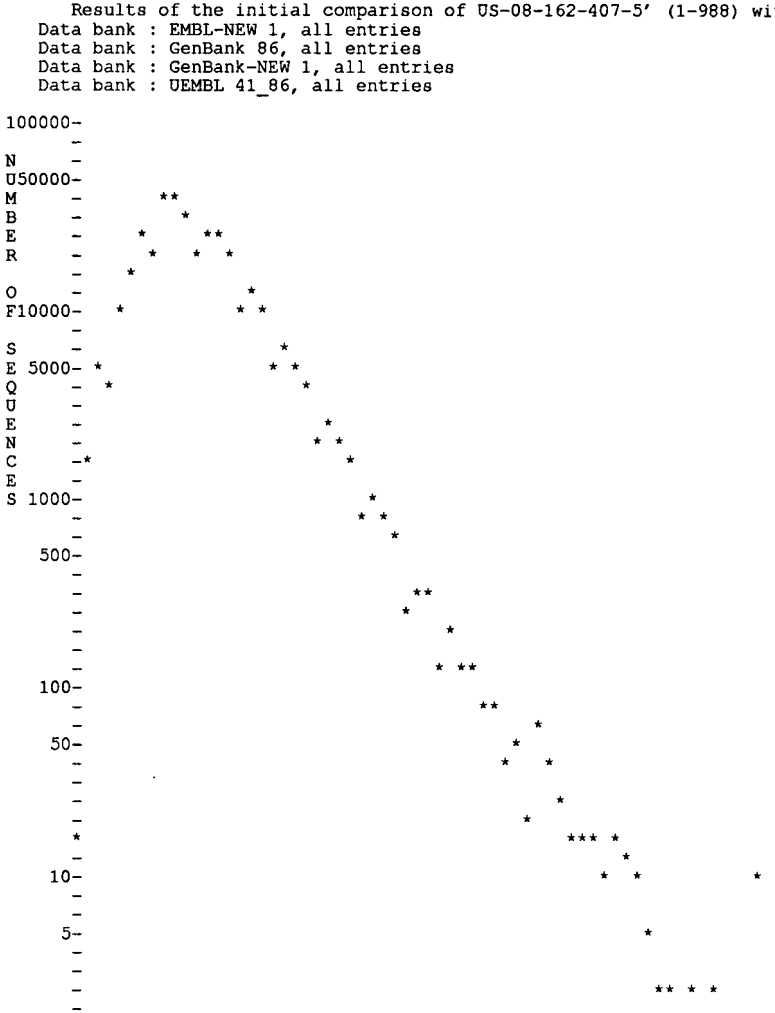


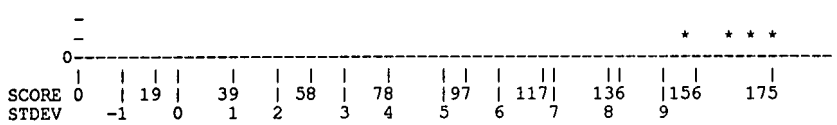
```
                     /note="NCBI gi: 603048"
                     /codon_start=1
                     /product="alpha-glucosidase"
                     /translation="MQQVPDSVHLTAPTAHTHWWRHAVIYQIYVRSFADSNGDGEGDL
                     NGIRERLPALVSLGVDAIWLTPFYVSPLADGGYDVADYRDVDPRFGTLADFDALLATA
                     HDLGLRVIIDIVPNHTSSAHRWFRDALAAGPGSPERDRYVFRPGRGENGELPPNNWQS
                     IFGGPAWTRVTEPDGTPGEWYLHLFDVEQPDLNWENPEVRAEFADILRFWLDRGVDGF
                     RIDVAHGMIKDPALPDIAEGQKADMLDGHTRLPYFDQDGVHEIYREWRAIVDSYPGER
                     ALVAEAWVENAERVARYLRPDELHQAFNFEYLTADWDAATLRAVVDRSLAANNAVGAP
                     TTWVLSNHDVTRHVTRFGGGAQGLARARAALLLMLALPGSVYLYQGEELGLPEVTDLP
                     EEALQDPTWKRSGYTERGRDGCRVPLPWEGDEPPFGFGCSAERSWLPVPAEWRSLTRE
                     VQERDPDSTLSLYKKALRLRRELLLPAEDALHWADAPQNVLAFRREPGFTCAVNFGAD
                     PVTLPFEGEVVLSSGPVEQDGTHLVLPGDTAVWLRQ"
BASE COUNT      342 a    725 c    703 g    323 t
ORIGIN

Initial Score    =    185  Optimized Score  =    450  Significance = 11.04
Residue Identity =    50%  Matches          =    541  Mismatches   =   391
Gaps             =    130  Conservative Substitutions               =     0

            X       10        20                30        40          50
            CGGCCGGAATTCCGGGGCCCCC--------GGCCGAAATGACAGTGCTG-GCGCCAGCCTGG
            || | |  | ||||||  | |        ||||||   |  |||  |  |||    |  |
   GCGCCATCGTCGACAGCTACCCCGGGGAGCGCGCGCTGGTGGCCGAGGCGTGGGTGGAGAACGCGGAACGCG
      910    X  920        930        940        950        960        970

         60        70        80        90        100               110        120
   AGCCCAACAACCTATCTCCTCCTGCTGCTGCTGCTGAGCTCGGGACT--CAGT--GGGACCCAGGACT-GCT
    | ||  | ||||  | || |     |||||  | | ||     |||   |||     |||   |||| |
   TGGCCCGCTACCT-CCGCCCCGACGAGCTGCACCAG-GCGTTCAACTTCGAGTACCTCACCGCAGACTGGGA
    980         990       1000      1010       1020       1030       1040

          130         140         150         160         170         180
   CCTTCCAACAC--AGCCCCAT-CTCCTCCG--ACTTCGCTGTCAA-AATCCG-TGAGCTGTCTGACTACCTG
   |  | |||| |   || || |  |   |||   | | || |  || ||  || || |    | ||| || ||
   CGCTGCAACGCTGCGCGCCGTGGTGGACCGCTCCCTGGCCGCGAACAACGCGGTGGGGGCGCCGACCACATG
   1050       1060       1070       1080       1090       1100       1110

     190        200        210        220        230        240        250
   CTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGC-CTCTGGCGGCT
     |   | |   |||  |||   | | | || || ||| | |     | || |||||||| | |  | ||||
   GGT---GCT-CTCCAACCAC---GACGT-CACCCGGCACGTCACCCGGTTCGGCGGGGGCGCGCAAG-GGCT
 1120          1130          1140       1150       1160       1170       1180

      260        270        280        290        300        310        320
   GGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGGAGC-
   ||  | ||||| |||| |    |   | ||| | | ||| | |||    ||  |||     | ||  || | |
   GGCGCGGGCAC-GCGCCG--CGCTCCTGCTGATG-CTGGCCCTG---CCGGGAT--CGGTCTACCTCTACCA
        1190          1200       1210       1220           1230       1240

   330        340        350        360        370             380        390
   GCGTGAACA-CGGAGA-TACACTTTGTCACCAAATGTGCC-TTTCAGCCCC------CCCCCAGCTGTCTTC
   | ||||  | | | || | | |   ||||||  || ||||     || |||        |||||  ||    )
   GGGTGAGGAGCTGGGACTTCCCGAGGTCACC-GATCTGCCGGAGGAGGCCCTGCAGGACCCCACGTGGAAGC
     1250       1260       1270       1280       1290       1300       1310

          400        410        420        430        440        450        460
   GCTTCGTCCAGACC-AACATCTCC-CGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTG
   ||| ||   | ||| | |    || || |   ||| ||   || | |   |    | ||||  | |||||
   GCTCCGGATACACCGAGCGGGGCCGCGACGGATGCCGGGTGCCGCTG-CCTTGGGAAGGCGACG-AGCCCCC
```

```
     1320       1330       1340       1350       1360       1370       1380

           470        480        490        500        510        520
    GATCACTCGCCAGAACTTCTC-CCGGTGC-CTG-GAGCTG-CAGTGTCAGCCCGACTCCTCAACCCTGCCAC
    | ||   |  | |  || |||  ||| || | |   |||| | ||| | | | || |       | |||  ||
    GTTC-GGCTTCGG--CTGCTCGGCGGAGCGCAGTTGGCTGCCGGTGCCCG-CGGAGTGGCGGTCGCTG--AC
      1390        1400       1410       1420       1430       1440       1450

   530        540        550        560        570        580        590
    CCCCATGGAGTCCCCG--GCCCCTGGAGGCCACAGCCCCGACAGC-----CCCGCAGCCCCCTCTGCTCCTCC
    ||    ||||  || |  || |     || | |  ||| | | ||    |  | || | |  | ||||| ||
    CC---GGGAGGTCCAGGAGCGCGACCCGGACTCGACCCTGTC-GCTGTACAAGAAGGCGCTGCGGCTCCGCC
           1460       1470       1480       1490       1500       1510

          600        610        620        630        640        650        660
    ---TACTGCTGCTGCCCG--TGGGCCTC-CTGCTGCTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGC
       |||||||||| ||||    || | | ||||    ||| |  |||  |   | | ||  || |      |
    GCGAACTGCTGCTCCCCGCAGAGGACGCGCTGCACTGGGCGGACGCCCCG---CAGAACGTGCTGGCTTTCC
1520       1530       1540       1550       1560          1570       1580

          670        680        690        700        710        720        730
    GGCGGAGGACACCCG-CCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAG
    | ||  ||| |||  |   |    | || ||||  |  ||  | | | |||  || |  |||||   | |||
    GCCG--GGA-ACCGGGTTTCACCTGCGC-GGTGAACTTCG-GCGCTGACCC--GGTCACGCTGCCGTTCGAG
 1590          1600       1610       1620       1630       1640       1650

          740        750        760        770        780        790
    CACTGACCTGGCCAAGGCC--TCATCCTGCGGAGCCTTAAACAACGCA-GTGAGACAGACATCTATCATC-C
      | ||  |||    | ||     ||| | |||||    |    |||||  |    | | |      || | |
    GGC-GAAGTGGTGCTGTCCAGCGGTCCGGTGGAGCAGGACGGCACGCACCTCGTGCTGCCTGGGGACACCGC
        1660       1670       1680       1690       1700       1710       1720

  800        810          820          830        840        850        860
    CATTTTACAG-GGGAGGA---TACTGA---GGCACACAGAGGGGAGTCACCAGCCAGAGGATGTATAGCCTG
      | |  | | || || |       |||   ||| |   | | |  | |||| | ||||||  ||    | ||
    GGTGTGGCTGCGGCAGTAGCGGGGTGAGTCGGCGCGTGGCGCGCTGACACC-GACAGAGG-GGTCCGGACTT
      1730       1740       1750       1760       1770       1780       1790

        870        880                 890        900        910        920
    GACACAGAGGAAGTTGG-----CTAGAGG----CCGGTCCCTTC-CTTGGGCCCCTCTCATTCCCTCCCCAG
      |   | ||  ||| |      ||  | |    ||||    ||| |||||      ||||||  | |  | ||
    TTC---GGGGGCGTTCGGACCCCTCTATGTCAACCGGAGATTTCACTTGG--ATTACTCATTTCGTTGCAAG
         1800       1810       1820       1830       1840         1850       1860

     930        940        950        960        970          980        X
    AATGGAGGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCAAC-TC-TGAACAAAGCCCCCG
    ||  |||  ||| |     |||| |    ||   |  |  | ||| ||     ||  || ||||
    AA-AGAGAAAACTC----TTCCAACTTTTGCAAAAAGTTGCTAACGTCCACGCCACCGCACCCGGAAGCCGC
          1870           1880       1890       1900       1910       1920

    GC
  1930
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq5inv.res made by  on Wed 5 Apr 95 20:39:41-PDT.
```



```
Query sequence being compared:US-08-162-407-5' (1-988)
Number of sequences searched:                313646
Number of scores above cutoff:                 4688

     Results of the initial comparison of US-08-162-407-5' (1-988) with:
   Data bank : EMBL-NEW 1, all entries
   Data bank : GenBank 86, all entries
   Data bank : GenBank-NEW 1, all entries
   Data bank : UEMBL 41_86, all entries
```





PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	30	30	13.31

Times:	CPU	Total Elapsed
	00:42:31.05	00:42:38.00

Number of residues: 288682119
Number of sequences searched: 313646
Number of scores above cutoff: 4688

Cut-off raised to 23.
Cut-off raised to 27.
Cut-off raised to 30.
Cut-off raised to 33.
Cut-off raised to 36.
Cut-off raised to 39.
Cut-off raised to 42.
Cut-off raised to 45.
Cut-off raised to 49.
Cut-off raised to 52.
Cut-off raised to 55.
Cut-off raised to 58.
Cut-off raised to 61.
Cut-off raised to 65.
Cut-off raised to 68.
Cut-off raised to 71.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame

**** 10 standard deviations above mean ****

1.	GORINVOLUC	Gorilla gorilla involucrin ge	1488	175	463	10.89	0
2.	HUMPKNB	Human mRNA for PKN.	2968	171	350	10.59	0
3.	HS4ULIR3	epstein-barr virus simple rep	1150	170	454	10.52	0
4.	U02454	Cloning vector pCMVEBNA, comp	5452	170	454	10.52	0
5.	U02428	Cloning vector pDR2, complete	10737	170	454	10.52	0
6.	U02455	Cloning vector rpDR2, complet	10850	170	454	10.52	0
7.	EBV	Epstein-Barr virus (EBV) geno	172281	170	454	10.52	0
8.	HS4B958RAJ	Epstein-Barr virus, artifactu	184113	170	454	10.52	0
9.	HS4B958RAJ	Epstein-Barr virus, artifactu	184113	170	454	10.52	0
10.	EBV	Epstein-Barr virus (EBV) geno	172281	170	454	10.52	0
11.	EBV	Epstein-Barr virus (EBV) geno	172281	170	454	10.52	0
12.	HUMLSPRO	Human lymphocyte surface prot	2622	167	464	10.29	0
13.	STMCARA	Streptomyces thermotolerans c	2393	164	448	10.07	0
		**** 9 standard deviations above mean ****					
14.	GORINVOLUB	Gorilla gorilla involucrin ge	1818	160	462	9.76	0
15.	GORINVOLUA	Gorilla gorilla involucrin ge	1908	160	462	9.76	0
16.	PPU12557	Pseudomonas putida PRS1 catBC	3686	155	452	9.39	0
17.	PP12557	Pseudomonas putida PRS1 catBC	3686	155	452	9.39	0
18.	STMCARB	Streptomyces thermotolerans g	2749	153	456	9.24	0
		**** 8 standard deviations above mean ****					
19.	SSRYR1B	S.scrofa RYR1 gene for ryanod	1546	148	418	8.86	0
20.	SCHCOADH	S.coelicolor genes for hydrox	5670	148	383	8.86	0
21.	ORAINVOL	Orang-utan involucrin gene, c	2508	147	460	8.79	0
22.	MASHCA	Mesocricetus auratus alpha-ca	32415	146	431	8.71	0
23.	HUMMMDBC	Human DNA from cosmid DNA MMD	68505	144	438	8.56	0
24.	HUMCACHNT	Human N-type calcium channel	7364	143	442	8.49	0
25.	HSCACHNT	Human N-type calcium channel	7364	143	442	8.49	0
26.	DROSEV3	D.virilis sevenless gene, exo	8779	143	451	8.49	0
27.	MUS5SRR5	Mouse gene for 5S rRNA.	1452	142	360	8.41	0
28.	MUS5SRR6	Mouse gene for 5S rRNA.	1476	142	359	8.41	0
29.	MMCRKING	M.musculus gene for mitochond	6031	142	400	8.41	0
30.	RNMTP1	R.norvegicus mtp1 mRNA.	2674	141	445	8.34	0
31.	HUMBCL2A	Human B-cell leukemia/lymphom	5086	140	320	8.26	0
32.	HSBCL2A	Human B-cell leukemia/lymphom	5086	140	320	8.26	0
33.	I08038	Sequence 1 from patent EP 025	5105	140	320	8.26	0
34.	HSREP10	Human beta-tubulin gene (5-be	8874	140	444	8.26	0
35.	CFCALCPH	Dog mRNA for calcyphosin.	1300	139	328	8.19	0
36.	CFCPHOS	Canis familiaris mRNA for cal	1304	139	327	8.19	0
37.	XELFIMC1X	Frog integumentary mucin (FIM	2161	139	317	8.19	0
38.	RATPKNA	Rat mRNA for PKN.	3035	139	386	8.19	0
39.	HUMIR3EB1	Human DNA with homology to EB	1400	138	349	8.11	0
40.	ECU08369	Escherichia coli phosphogluco	3176	138	433	8.11	0
41.	ECU08369	Escherichia coli K12 phosphog	3176	138	433	8.11	0
42.	EC08369	Escherichia coli phosphogluco	3176	138	433	8.11	0
43.	DSANTPE1	D.subobscura antennapedia hom	2016	137	422	8.04	0
44.	HUMGFB	Human basic fibroblast growth	6757	137	381	8.04	0
45.	HSGFB	Human basic fibroblast growth	6757	137	381	8.04	0

1. US-08-162-407-5' (1-988)
   GORINVOLUC Gorilla gorilla involucrin gene small allele, comp

LOCUS	GORINVOLUC 1488 bp ds-DNA PRI 07-SEP-1994
DEFINITION	Gorilla gorilla involucrin gene small allele, complete cds.
ACCESSION	M23605 J04499
KEYWORDS	epidermal protein; involucrin.
SOURCE	Gorilla gorilla (individual_isolate Gorilla K) vagina DNA.
ORGANISM	Gorilla gorilla
	Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;

```
            Eutheria; Primates; Haplorhini; Catarrhini; Pongidae.
  REFERENCE 1  (bases 1 to 1488)
    AUTHORS   Teumer,J. and Green,H.
    TITLE     Divergent evolution of part of the involucrin gene in the
              hominoids: Unique intragenic duplications in the gorilla and human
    JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86, 1283-1286 (1989)
    STANDARD  full automatic
  COMMENT     Draft entry and computer readable copy of sequence [1] kindly
              provided by J.K. Teumer (03-31-89).

              NCBI gi: 340979
  FEATURES             Location/Qualifiers
       source          1..1488
                       /organism="Gorilla gorilla"
                       /isolate="Gorilla K"
                       /cell_type="keratinocyte"
                       /sequenced_mol="DNA"
                       /clone="pIGorH5.8"
                       /tissue_type="vagina"
       CDS             1..1488
                       /note="NCBI gi: 536829"
                       /codon_start=1
                       /product="involucrin"
                       /translation="MSLQHTLPVTLSPALSQELLKTVPPPVNTQQEQMKPPTPLPPPC
                       QKVPVKLPVEVPSKQEEKHMTAVKGLPEQECEQQQQEPQEQEPQQQHWEQHEEHQKAE
                       NPEQQLKQEKAQRDQQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELPEQRE
                       GELKHLEQQKGQLELPEQQEGQLELPEQQEGQLKHLEQQEGQLELPEQQEGQLELPEQ
                       QEGQLKHLEQQEGQLELPEQQEGQLKHLEQQEGQLELPEQQEGQLELPEQQEGQLKHL
                       EQQEGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLEHLVQQEGQLEQQEGQVEHLE
                       QQVGQLKHLEEQEGQLKHLEQQQGQLGVPEQQVGQPNNLEQEEKQLELPEQQEGQLKH
                       LEKQEAQLELPEQQVGQPKHLEQQEKQLEHPEQQDGQLKHLEQQEGQLKDLEQQKGQL
                       EQPVFAPAPGQVQDIQPALPTKGELLLPVEHQQQKQEVQWPPKHK"
  BASE COUNT      445 a    383 c    515 g    145 t
  ORIGIN

Initial Score    =    175  Optimized Score  =    463  Significance = 10.89
Residue Identity =    52%  Matches          =    556  Mismatches   =   366
Gaps             =    141  Conservative Substitutions             =     0

            X        10        20        30        40        50
            CGGGGGCT-TTGTTCAGAGTTGGGTAAATGGGGCCGGTGCTGGA--TTCTGGCGTTGCCTCC
              |  | |||  |    |||| || | || ||||   |  |   ||    |||  |  ||  ||
   GCAACTGGATCAAGAGCTAGTCAAGAGAGATGAGCAACTGGGAATGAAGAAAGAGCAACTGTTGGAGCTCCC
     400     X 410       420       430       440       450       460

   60         70        80        90        100       110       120
    A-TTCTG-GGGAGGGAATGAG-AGGGGCCCAAGGA--AG--GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
    |   | | ||||||||  |||  |  || |  |||  ||  | |  |||  || |  | || ||   |
    AGAGCAGCGGGAGGGA--GAGCTGAAGCACCTGGAGCAGCAGAAGGGGCAGCTGG--AGCTCCCAGAG-CAG
    470       480       490       500       510       520       530

       130       140       150       160       170       180       190
    CAGGCTATACATCCTCTGGCTGGTGACTCCC--CTCTG--TGTGCCTCAGTATCCTCCCCTGTAAAATGGGA
    ||||     ||         ||||| | |||||     | |   | |  |||       | |||| |  |  | |
    CAGGAGGGGCA-------GCTGGAG-CTCCCAGAGCAGCAGGAGGGACAGCTGAAGCACCTGGAGCA--GCA
       540              550       560       570       580       590

         200       210       220       230       240       250       260
    TGATAGATGTCTG--TCTCACTGCGTTGTTTAAGGCTCCGCAGGATGAGGCCTTGGCCAGGTCAGTGCTCCA
     ||  |    |||   ||| | | |  |    |||  | || ||| |   ||  |  |||  ||| |    |
```



```
   GGAGGGGCAGCTGGAGCTCCCAGAGCAGCAGGAGGGGCAGCTGGA-GCTCCC-AGAGCAG--CAG-GAGGGA
    600       610       620       630       640       650       660

         270       280       290       300       310       320       330
   CAAGCAGCAGGTCCTGGGGACTGGGGACGGGGGGCACCTGCTCCCCAGGGC----GGGGTGTCCTCCGCCGC
   | ||| | ||  ||| | |     ||| ||| ||     ||| |||| ||    || | | |  | |  ||
   C-AGCTGAAGCACCTAGAGCAGCAGGAGGGGCAGCTGGAGCT-CCCAGAGCAGCAGGAGGGACAGCTGAAGC
        670       680       690       700       710       720       730

                340       350       360       370       380             390
   GTCCTCTGC--CAGTGCAGGCACCAGGCAGCGGCC--AGCAGCAGGAGGCCCACGGGCAGC------AGCAG
     |    ||  |||    |||| | || |||  ||  ||||||||||||  ||   | |||      |||||
   ACCTGGAGCAGCAGGAGGGGCAGCTGG-AGCTCCCAGAGCAGCAGGAGGGGCAGCTGGAGCTCCCAGAGCAG
       740       750       760       770       780       790       800

            400       410       420       430       440       450       460
   TAGGAGG---AGCAGAGGGGGCT-GCGGGGCTGTCGGGGCTGTGGCCTCCAGGGGCCGGGGACTCCATGGGG
    ||||||   ||| || |   || | |  || |  ||||| |    ||  ||   || | | |  ||  ||
   CAGGAGGGACAGCTGAAGCACCTAGAGCAGCAGGAGGGGCAG----CTGGAGCTCCCAGAG-CAGCA--GGA
      810       820       830       840           850       860

           470       480           490       500       510       520
   GTGGCAGGGTTGAGGAGTCGG----GCTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
   | | || | |  || |   ||     || ||     ||||||  | || ||  ||| ||  | || | |  | |
   GGGACA-GCTGAAGCACCTGGAGCAGCAGAAGGGGCAGCTGGA-GCTCCCAGAGCAG--CAGGAGGGGCAGC
870        880       890       900       910       920       930

 530       540       550       560       570       580       590       600
   CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
     | || || |   || |||       |  || |  ||| || | || ||||  |  ||  ||||  |   |
   TGGAGCATCTGGTGCAGCAG------GAGGGGCAGCTGGAGCA-GCAGGAGGGGCAGG--TGGAGCACCTGG
   940       950             960       970       980       990      1000

           610       620       630       640       650       660
   AGACAGCTGG-GGGGGGGCTGAAAGGCACAT-TTGGTG-ACAAAGTGTATCTCCGTGTTCACGCGCTCCAGC
   || |||| || |||| |||||| |||| |   || | |  | | | | ||       ||  | ||  ||||
   AG-CAGCAGGTGGGGCAGCTGAA--GCACCTAGAGGAGCAGGAGGGGCAACTGAAGCATCTGGAGCAGCAGC
          1010      1020      1030      1040      1050      1060

 670       680       690       700       710       720       730
   AAGCCTTGCA--TCTTGGACCCAGCGACAGTCTTGAGCCGC-TCCATCCAGCGCTGTGC--CAGGACCAGCC
   | |    |||  |   || |||| | ||| |  | |  ||  |||  || |   | ||  |||| |||
   AGG---GGCAGTTGGGGGTCCCAGAG-CAG-CAGGTGGGGCAGCCAAACAACCTGGAGCAGGAGGAGAAGCA
1070         1080      1090      1100      1110      1120      1130

    740       750       760       770       780       790
   GCCAGAGGCCCC--CGCA----GAGCTCCTCGT------CCTGCAG--GTTGGAGGCCACGGTGACTGGGTA
   ||  |||  |||   |||    |||  ||   |      |||| ||  |  ||||| ||| |   ||||
   GCTGGAGCTCCCAGAGCAGCAAGAGGGCCAGCTGAAGCACCTGGAGAAGCAGGAGG-CACAG---CTGGAGC
   1140      1150      1160      1170      1180      1190          1200

      800       810       820       830       840       850       860
   ATCTTGA--AGCAGGTAGTCAGACAGCTCACGGATTTTGA-CAGCGAAGTCGGAGGAGATGGGGCTGTGTTG
     |  ||  |||||||||   ||||||  | | |  | || ||||  ||     || || | | ||
   TCCCAGAGCAGCAGGTAG---GACAGCCAAAGCACCTGGAGCAGC--AGGAAAAGCAGCTAGAGC--ACCCA
     1210      1220      1230      1240      1250      1260

       870       880       890       900       910       920       930
   GAAGGAGCAGTCCTGGGTCCCACTGAGTCCCGAGC-TCAGCAGCAGCAGGAGGAGATAG-GTTGTTG-G
   || | |||||  | |    | |     ||  ||||  ||| ||  |||||  | ||||  | | |  |  |
```

```
   GA-GCAGCAGGACGGACAACTAAAACATCTGGAGCAGCAGGAGGGGCAGCTGAAGGACCTGGAGCAGCAGAA
1270       1280       1290       1300       1310       1320       1330       1340

        940        950        960                970        980        X
   GCTCCAGGCTGGCGCCAGCACTGTCATTTC-----GGCCGGG---GGCCCCGGA-ATTCCGGCCG
   |   || |||||  | ||||  ||||  || |       || ||   || ||  || ||||   | |
   GGGGCA-GCTGGAG-CAGC-CTGTGTTTGCCCCAGCTCCAGGCCAGGTCCAAGACATTCAACCAGCCCTGCC
         1350       1360       1370       1380       1390       1400 X

   CAC
1410
```

2. US-08-162-407-5' (1-988)
   HUMPKNB Human mRNA for PKN.

```
LOCUS       HUMPKNB      2968 bp ss-mRNA            PRI       12-MAY-1994
DEFINITION  Human mRNA for PKN.
ACCESSION   D26181
KEYWORDS    PKN.
SOURCE      Homo sapiens (library: lambda ZAPII) hippocampus cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 2968)
  AUTHORS   Ono,Y. and Mukai,H.
  TITLE     A novel protein kinase with leucine zipper-like sequences: Its
            catalytic domain is highly homologous to that of protein kinase C
  JOURNAL   Biochem. Biophys. Res. Commun. 199, 897-904 (1994)
  STANDARD  full automatic
COMMENT     Submitted (22-Dec-1993) to DDBJ by: Yoshitaka Ono
            Dept. of Biochemistry
            Faculty of Science
            Kobe University
            1-1 Rokkodai-cho, Nada-ku
            Kobe 657
            Japan
            Phone: 078-881-1212 x4450
            Email: yono@niguts.nig.ac.jp
            Fax:   078-801-1740.

            NCBI gi: 473712
FEATURES             Location/Qualifiers
     source          1..2968
                     /organism="Homo sapiens"
                     /sequenced_mol="cDNA to mRNA"
                     /tissue_type="hippocampus"
                     /clone_lib="lambda ZAPII"
BASE COUNT      587 a    928 c    961 g    492 t
ORIGIN
```

```
Initial Score    =    171  Optimized Score  =    350  Significance = 10.59
Residue Identity =    52%  Matches          =    416  Mismatches   =   271
Gaps             =     99  Conservative Substitutions               =     0

      250    X  260        270        280        290        300        310
   GGCCAGGTCAGTGCTCCACAAGCAGCAGGTCCTGGGGACTGG-GGACGGGGGGCACCTGCTCCCCAGGGCGG
              |   ||| |  ||||    ||| ||   | ||  |||  ||    || ||  | | | || |
              GAATTCC-CGCGCAGAGACTCCAGGTCGCAGGTCGACATGG----CCAGCGACGCCGTGCAG
              X        10         20         30         40            50
```



```
   320        330        340        350        360        370        380
GGTGTCCTCCGCCGC--GTCCTCTGCCAGTGCAG--GCACCAGGCAGCGGCCAGCAGCAGGAGGCCCACGGG
 |||  |  ||| ||  |||| |||| || ||||  |  || ||| | || ||| | | || ||||| ||||
AGTGAGCCTCGCAGCTGGTCC-CTGCTAGAGCAGCTGGGCCTGGCCG-GGGCAG-ACCTGGCGGCCCCCGGG
 60         70         80         90         100        110        120

    390        400        410        420        430        440        450
CAGCAGCAGTAGGAGGAGCAGAGGGGGCTGCGGGGCTGTCGGGGCTGTGGCCTCCAGGGGCCGGGGACTCCA
   ||||||  ||  ||||| |  | ||  | | ||||| ||| |       || | ||| || |  | ||  |
GTACAGCAGCAGCTGGAGCTGGAGCGG--GAGCGGCTG-CGGCGGGAAATCCGCAAGGAGCTGAAG-CTGAA
 130        140        150        160        170        180        190

  460        470        480        490                500        510
TGGGGGTGGCAGGGTTGAGGAGTCGGGCTGACACT--------GCAGC-TCCAGGCACCGGGAGAAGTTCTG-
 | |||| || | |      | | ||||  ||||        || ||  || || ||  || | |||  | |||
GGAGGGT-GCTGAGAACCTGCGGCGGGCCACCACTGACCTGGGCCGCAGCCTGGGCCCCGTAGAGCTGCTGC
   200        210        220        230        240        250        260

520        530        540        550        560        570        580
-GCGAGTGATCCAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAG--GC-GGGAGATGTTGG
 ||| | | |||    ||  |  | | ||| |||||  |  ||  |  |||||||||  ||  |   | || |
TGCG-GGGCTCC-TCGCGCCGCCTCGACCTGCTGCAC-CAG--CAGCTGCAGGAGCTGCACGCCCACGTGGT
   270        280        290        300        310        320        330

 590        600        610        620        630        640        650
TCTGGACGAAGCGAAGACAGCTGGGGGGGGGCTGAAAGGCACATTTGGTGACAAAGTGTATCTCCGTGTTCA
 ||   |||  ||  | | |      |  |||    || | |   |||||     | ||  | ||   | |
GCTTCCCGACCCGGCGGCCAC-CCACGATGGCCCCCAGTCCC--CTGGTG----CGGGTGGCCCCACCTGCT
     340        350        360        370            380        390

660        670        680        690            700        710        720
CGCGCTCCAGC--AAGCCTTGCATCTTGGACCCAGCGA--CAGT-----CTTGAGCCGCTCCATCCAGCGCT
||  | ||| |   ||||  || |   || ||  | ||  |||       ||||||  |     |  |||
CGGCCACCAACCTGAGCC--GCGTGGCGGGCCTGGAGAAGCAGTTGGCCATTGAGCTG-AAGGTGAAGCAGG
400        410        420        430        440        450        460

      730        740        750            760        770        780
GTGCCAGGACCAGCCGCCAGAGGCCCCCGCA----GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGACT
| ||   || ||    |||||  |  | |||     ||| ||  |  | | | | ||  |   ||   | |
GGGCGGAGAACATGATCCAGA-CCTACAGCAATGGCAGCACCAAGGACCGGAAGCTGCTGCTGAC--AGCCC
 470        480        490        500        510        520        530

790        800        810        820        830        840        850
 GGGTAATCTTGAAGCAGGTAGTCAGACAGCTCACGGATTTTGACAGCGAAGTCGGAGGAGA-----TGGGGC
  |   || ||| ||  |  ||| ||||    ||  |||  ||||| |  |  || | |  |         ||||
 AGCAGATGTTGCAG--GACAGTAAGAC----CA-AGAT--TGACATC--ATCCGCATGCAACTCCGCCGGGC
   540        550        560            570        580        590

 860        870        880        890        900        910            920
TGTGTTGGAAGGAGCAGTCCTGG-GTCCCACTGAGTCCC-GAGCTCAGCAGCAGCAG-CAGGA----GGAGA
  ||  ||  ||  |||  |||| |  |||   || ||| ||   || |    | || |  ||    || |
GCTGCAGGCCGG-CCAG--CTGGAGAACCAGGCAGCCCCGGATGACACCCAAGGGAGTCCTGACCTGGGGGC
 600        610        620        630        640        650        660

      930            940        950        960        970        980
TAGGTTGTTGGGC-TCCA--GGCTGGCGCCAGCACTGTC-ATTTCGGC---CGGGGGCCCCGGAATTCC--G
|  |  | || || || |   ||| ||| || ||||  | | |   ||   ||| ||||  ||  | ||  |
TGTGGAGCTGCGCATCGAAGAGCT-GCGGCACCACTTCCGAGTGGAGCACGCGGTGGCCGAGG-GTGCCAAG
 670        680        690        700        710        720        730
```

```
      X
    GCCG
      ||
    AACGTACTGCGCCT
      740
```

3. US-08-162-407-5' (1-988)
   HS4ULIR3 epstein-barr virus simple repeat array (ir3).

```
LOCUS       HS4ULIR3     1150 bp ds-DNA             VRL       15-JUN-1990
DEFINITION  epstein-barr virus simple repeat array (ir3).
ACCESSION   J02079
KEYWORDS    repeat region.
SOURCE      ebv (epstein barr virus) from human.
  ORGANISM  Epstein-Barr virus
            Viridae; ds-DNA enveloped viruses; Herpesviridae;
            Gammaherpesvirinae.
REFERENCE   1  (bases 1 to 1150)
  AUTHORS   Heller,M., van Santen,V.L. and Keiff,E.
  TITLE     simple repeat sequence in epstein-barr virus dna is transcribed in
            latent and productive infections
  JOURNAL   J. Virol. 44, 311-320 (1982)
  STANDARD  full automatic
COMMENT     NCBI gi: 330445
FEATURES             Location/Qualifiers
     source          1..1150
                     /organism="Epstein-Barr virus"
BASE COUNT      302 a    169 c    633 g     46 t
ORIGIN
```

Initial Score = 170 Optimized Score = 454 Significance = 10.52
Residue Identity = 50% Matches = 529 Mismatches = 418
Gaps = 105 Conservative Substitutions = 0

```
             X       10          20        30        40        50
             CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
             ||| |||  ||    ||| ||  | |    ||    | ||   ||   || | |    |
   CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
        80 X      90        100       110       120       130       140

   60        70         80        90          100       110       120
   CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
    |    | ||||| | | ||  ||| ||||  ||  |    ||| | |   | || ||       |  |||||
   GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
       150       160       170       180       190       200       210

      130       140       150       160       170       180       190
   CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
   | |   | ||  ||      |     | |  | | ||  || |  ||   ||  |  || || | | | | ||
   CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
         220       230       240       250       260       270       280

      200       210       220       230       240       250       260
   -TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
     ||  |   |     |   | |    |||  |    |||||| || || | |  ||||   ||  ||
   GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
          290       300       310       320       330       340       350
```



```
          270       280       290       300       310       320
   CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
    |   |||| |||  | |  ||| | ||    ||| | ||||||  |     ||| | || | | |   |
   GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
          360       370       380       390       400       410

    330       340       350       360       370       380       390
   CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
     ||       ||  | |||||  |  |||  || |  ||||| ||||||  |   | ||||| ||||| ||
   GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
    420       430       440       450       460       470       480

      400       410       420       430       440       450       460
   GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
   |||| ||| ||||||   | |||||  |  |  ||  | ||    |||| || || |||    || | ||||
   GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
     490       500       510       520       530       540       550

          470       480       490       500       510       520
   GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
    | ||  ||  ||||   ||  ||| | |    ||||   ||||    |||  | |  | || | |  |
   AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
     560       570       580       590       600       610       620

  530       540       550       560       570       580       590       600
   CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
     ||||   || | ||  ||| |   |  || |    |||||||| ||| | |  || | ||| |    |
   AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
   630       640       650       660       670       680       690

          610       620       630       640       650       660
   AGACAGCTGGGG-GGGGGC-TGAAAGGCA-----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
   || |||  ||||  || ||  ||  ||||     ||   || | ||  | | |      |    | | ||
   AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
     700       710       720       730       740       750       760

    670       680       690       700       710       720       730
   AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
   ||||  |    |||    || |  |||  ||  | |  |||  ||   | |  | | | | | | |||||     |
   AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
     770       780       790       800       810       820       830

        740       750       760       770       780       790
   AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA----CTGGGTAATCTT
   ||  ||  ||||      ||| |||    |  |  | | |||  | |||  | ||  |     | ||      |
   AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGAGGGCAG
   840       850       860       870       880       890       900

800       810       820       830       840       850       860
   GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
   || |  || ||   |||    |  || |||     | || | | ||  ||||| |  || || |    || |||
   GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
       910       920       930       940       950       960       970

  870       880       890       900       910       920       930
   AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
   |||||     ||   | | | ||  ||||    ||||  |  || || |  ||  |||| | |||   || |
   AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
        980       990      1000      1010      1020      1030
```

```
  940       950       960           970        980       X
    GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
    | | | |   | | ||  |   || | |||    || |  | ||   || |  |
    GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
1040      1050      1060      1070      1080      1090 X    1100
```

4. US-08-162-407-5' (1-988)
   U02454 Cloning vector pCMVEBNA, complete sequence.

```
LOCUS       U02454       5452 bp    DNA   circular  SYN       08-NOV-1993
DEFINITION  Cloning vector pCMVEBNA, complete sequence.
ACCESSION   U02454
KEYWORDS    .
SOURCE      Cloning vector pCMVEBNA.
  ORGANISM  Cloning vector pCMVEBNA
            Artificial sequences; Cloning vector.
REFERENCE   1  (bases 1 to 5452)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.1
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 5452)
  AUTHORS   Swirski,R.A., Van Den Berg,D., Murphy,A.J., Lambert,C.M.,
            Friedberg,E.C. and Schimke,R.T.
  TITLE     Improvements in the Epstein-Barr-based shuttle vector system for
            direct cloning in human tissue culture cells
  JOURNAL   Methods: A Companion to Methods in Enzymology 4, 133-142 (1992)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 5452)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            4030 Fabian Way, Palo Alto, CA 94303, USA
  STANDARD  full automatic
COMMENT     This vector can be obtained from CLONTECH Laboratories, Inc., 4030
            Fabian Way, Palo Alto, CA 94303, USA. To place an order call (415)
            424-8222 or (800) 662-2566, extension 1. International customers,
            please contact your local distributor.  For technical information,
            call (415) 424- 8222 or (800) 662-2566, extension 3.
            This sequence has been compiled from information in the sequence
            databases, published literature and other sources; this vector has
            not been completely sequenced. If you suspect there is an error in
            this sequence, please contact CLONTECH's Technical Service
            Department at (415) 424-8222 or (800) 662-2566, extension 3 or
            E-mail CLONTECH@BIOTECHNET.COM.

            NCBI gi: 413820
FEATURES              Location/Qualifiers
     source           1..5452
                      /organism="Cloning vector pCMVEBNA"
BASE COUNT     1362 a   1246 c   1736 g   1108 t
ORIGIN
```

```
Initial Score     =    170  Optimized Score  =    454  Significance = 10.52
Residue Identity  =    50%  Matches          =    529  Mismatches   =   418
Gaps              =    105  Conservative Substitutions               =     0
```

```
             X       10          20        30        40        50
             CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
```



```
                 |||  |||  ||    |||  ||  | |    ||    |  ||   ||    ||  |  |    |
CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
   870    X  880       890       900       910       920       930

60        70          80        90          100       110       120
CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
 |    | |||||  | | ||  |||  ||||  ||  |    |||  | |   | ||  ||       |  |||||
GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
 940       950       960       970       980       990      1000

   130       140       150       160       170       180       190
CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
| |   | ||  ||      |     | |  | | ||  || |  ||   ||  |  || || | | | | ||
CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
1010          1020      1030      1040      1050      1060      1070

   200       210       220          230       240       250          260
-TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
  ||  |   |      |   | |    |||  |    ||||||  || || |  |  ||||    ||  ||
GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
   1080      1090        1100      1110      1120      1130      1140

        270         280           290       300       310       320
CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
 |   |||| |||  | |  ||| | ||    ||| | ||||||  |     ||| | || | | |    |
GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
   1150      1160      1170      1180      1190          1200

  330       340       350       360       370          380       390
CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
  ||        ||  | |||||  |  |||  || |  ||||| ||||||  |   | |||||| ||||| ||
GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
1210      1220      1230      1240      1250      1260      1270

   400       410       420       430       440       450       460
GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
|||| ||| ||||||   | |||||  |  |  ||  | ||    |||| || || |||    || | ||||
GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
1280      1290      1300      1310      1320      1330      1340

        470         480       490       500       510       520
GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
 | ||  ||  ||||   ||  ||| | |    ||||  ||||    ||| | |  | || | |  |
AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGGCAGGAGCAGGAGGGGCAGG
1350      1360      1370      1380      1390      1400      1410      1420

 530       540       550       560       570       580       590       600
CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
  ||||   || | ||  ||| |   |  || |    |||||||| |||  | |  || | ||| |    |
AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
      1430      1440        1450      1460      1470      1480

        610         620           630       640       650       660
AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
|| |||  ||||  || ||  ||  ||||     ||   || | ||  | | |     |    | | ||
AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
1490      1500      1510      1520      1530      1540      1550

  670       680       690       700       710       720       730
AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
```

```
     |||| |    |||    || |   |||  ||  | |  |||  ||   | |  | |  | | |||||    |
     AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
1560       1570       1580       1590       1600       1610       1620       1630

          740        750        760        770        780            790
     AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA----CTGGGTAATCTT
     ||  ||  ||||      ||| |||   |  |  | | ||| | |||  | ||  |     | ||     |
     AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
          1640       1650       1660       1670       1680       1690

800        810          820        830        840        850        860
     GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
     || |  || ||   |||   |  || |||    | || | | ||  ||||| |  || || |    || |||
     GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
1700         1710       1720          1730       1740       1750       1760

 870        880        890        900        910        920        930
     AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
     |||||      ||   | | | ||  ||||    ||||  |  || || |  ||  |||| | |||   || |
     AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
      1770       1780          1790       1800       1810       1820       1830

940        950        960             970        980        X
     GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
     | | | |   | | ||  |   || | |||    || |  | ||   || | |
     GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
         1840       1850       1860       1870       1880     X 1890
```

5. US-08-162-407-5' (1-988)
   U02428 Cloning vector pDR2, complete sequence.

```
LOCUS       U02428      10737 bp    DNA   circular  SYN       08-NOV-1993
DEFINITION  Cloning vector pDR2, complete sequence.
ACCESSION   U02428
KEYWORDS    .
SOURCE      Cloning vector pDR2.
  ORGANISM  Cloning vector pDR2
            Artificial sequences; Cloning vector.
REFERENCE   1  (bases 1 to 10737)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.1
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 10737)
  AUTHORS   Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
  TITLE     cDNA expression cloning in human cells using the plambdaDR2
            episomal vector system
  JOURNAL   Methods:A Companion to Methods in Enzymology 4, 111-131 (1992)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 10737)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            4030 Fabian Way, Palo Alto, CA 94303, USA
  STANDARD  full automatic
COMMENT     This vector can be obtained from CLONTECH Laboratories, Inc., 4030
            Fabian Way, Palo Alto, CA 94303, USA. To place an order call (415)
            424-8222 or (800) 662-2566, extension 1. International customers,
            please contact your local distributor.  For technical information,
```



```
            call (415) 424- 8222 or (800) 662-2566, extension 3. This sequence
            was compiled by Andrew Murphy and revised at CLONTECH; this vector
            has not been completely sequenced. If you suspect there is an error
            in this sequence, please contact CLONTECH's Technical Service
            Department at (415) 424-8222 or (800) 662-2566, extension 3 or
            E-mail CLONTECH@BIOTECHNET.COM.

            NCBI gi: 413794
FEATURES             Location/Qualifiers
     source          1..10737
                     /organism="Cloning vector pDR2"
BASE COUNT     2686 a    2596 c    3033 g    2422 t
ORIGIN
```

```
Initial Score    =    170  Optimized Score  =    454  Significance = 10.52
Residue Identity =    50%  Matches          =    529  Mismatches   =   418
Gaps             =    105  Conservative Substitutions             =     0

              X       10          20          30          40          50
              CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
              ||| |||  ||    ||| ||  | |    ||    |  ||   ||   || | |    |
    CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
  6400       6410       6420       6430       6440       6450       6460

   60         70          80          90          100         110         120
   CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
    |    | ||||| | | ||  ||| ||||  ||  |    ||| | |   | || ||       |  |||||
   GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
6470       6480       6490       6500       6510       6520       6530       6540

     130        140         150         160         170         180         190
   CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
   | |   | ||  ||      |    | |  | | ||  || |  ||   ||  |  || || | | | | ||
   CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
            6550       6560       6570       6580       6590       6600

    200        210         220         230         240         250         260
   -TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
     ||  |   |     |   | |    |||  |   ||||||  || || |  |  ||||   ||  ||
   GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
   6610       6620        6630       6640       6650       6660       6670

         270         280         290         300         310         320
   CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
    |   |||| |||  | |  ||| | ||    ||| | ||||||   |     ||| | || | | |    |
   GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
   6680       6690       6700       6710       6720            6730       6740

    330        340         350         360         370         380         390
   CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
     ||       ||  | |||||  |  |||  || |  ||||| ||||||  |   | ||||||  ||||| ||
   GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
      6750       6760       6770        6780       6790       6800       6810

     400        410         420         430         440         450         460
   GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
   ||||  ||| ||||||   | |||||  |  |  ||  | ||    |||| || || |||    || | ||||
   GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
       6820       6830       6840       6850        6860       6870       6880
```

```
        470         480         490         500         510         520
GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
  | ||  ||  ||||    ||  ||| | |      ||||   ||||    |||  | |  | || | |  |
AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
     6890        6900        6910        6920        6930        6940        6950

530         540         550         560         570         580         590         600
CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
  ||||   || | ||  ||| |   |  || |    |||||||| |||  | |  || | ||| |    |
AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
   6960        6970        6980        6990        7000        7010        7020

        610         620         630         640         650         660
AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
|| |||  ||||  || ||  ||  ||||     ||   || | ||  | | |      |     | | ||
AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
     7030        7040        7050        7060        7070        7080        7090

 670         680         690         700         710         720         730
AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
||||  |    |||    || |   |||  ||  | |  |||  ||   | |  | |  | |  | | |||||    |
AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
     7100        7110        7120        7130        7140        7150        7160

        740         750         760         770         780         790
AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA----CTGGGTAATCTT
||  ||  ||||     ||| |||   |  |  | | |||  | |||  | ||  |     | ||     |
AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
   7170        7180        7190        7200        7210        7220        7230

800         810         820         830         840         850         860
  GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
  || |  || ||   |||   |  || |||    | || | | ||  ||||| |  || || |    || |||
  GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
        7240        7250        7260        7270        7280        7290

870         880         890         900         910         920         930
  AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
  |||||     ||   | | | ||  ||||    ||||  |  || || |  ||  |||| | |||   || |
  AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
7300        7310        7320        7330        7340        7350        7360

940         950         960         970         980         X
  GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
  | | | |   | | ||  |   || | |||    || |  | ||   || | |
  GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
   7370        7380        7390        7400        7410        7420
```

6. US-08-162-407-5' (1-988)
   U02455 Cloning vector rpDR2, complete sequence.

```
LOCUS       U02455      10850 bp    DNA   circular  SYN       08-NOV-1993
DEFINITION  Cloning vector rpDR2, complete sequence.
ACCESSION   U02455
KEYWORDS    .
SOURCE      Cloning vector rpDR2.
  ORGANISM  Cloning vector rpDR2
            Artificial sequences; Cloning vector.
REFERENCE   1  (bases 1 to 10850)
```



```
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.1
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 10850)
  AUTHORS   Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
  TITLE     cDNA expression cloning in human cells using the plambdaDR2
            episomal vector system
  JOURNAL   Methods: A Companion to Methods in Enzymology 4, 111-131 (1992)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 10850)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            4030 Fabian Way, Palo Alto, CA 94303, USA
  STANDARD  full automatic
COMMENT     Lambda DR2 can be obtained from CLONTECH Laboratories, Inc., 4030
            Fabian Way, Palo Alto, CA 94303, USA. To place an order call (415)
            424-8222 or (800) 662-2566, extension 1. International customers,
            please contact your local distributor.  For technical information,
            call (415) 424- 8222 or (800) 662-2566, extension 3.
            This sequence was compiled by Andrew Murphy and revised at
            CLONTECH; this vector has not been completely sequenced. If you
            suspect there is an error in this sequence, please contact
            CLONTECH's Technical Service Department at (415) 424-8222 or (800)
            662-2566, extension 3 or E-mail CLONTECH@BIOTECHNET.COM.

            NCBI gi: 413821
FEATURES             Location/Qualifiers
     source          1..10850
                     /organism="Cloning vector rpDR2"
                     /note="plasmid released from lambda DR2"
BASE COUNT     2723 a   2615 c   3053 g   2459 t
ORIGIN
```

```
Initial Score    =    170  Optimized Score  =    454  Significance = 10.52
Residue Identity =    50%  Matches          =    529  Mismatches   =   418
Gaps             =    105  Conservative Substitutions              =     0

             X       10          20          30          40          50
             CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
             ||| |||  ||    ||| ||  | |    ||    |  ||   ||    || |  |    |
   CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
         6520        6530        6540        6550        6560        6570        6580

   60          70          80          90          100         110         120
   CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
    |    | ||||| | | ||  ||| |||| ||  |    ||| | |   | || ||      |  |||||
   GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
       6590        6600        6610        6620        6630        6640        6650

      130         140         150         160         170         180         190
   CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
   | |   | ||  ||      |     | |  | | ||  || | |  ||   ||  |  || || | | | ||
   CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
         6660        6670        6680        6690        6700        6710

      200         210         220         230         240         250         260
   -TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
     ||  |   |     |   | |    |||  |    |||||| || || |  |  ||||   ||  ||
```

```
   GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
6720       6730       6740       6750       6760       6770       6780

          270        280        290        300        310        320
   CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
   GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
6790       6800       6810       6820       6830       6840       6850

    330        340        350        360        370        380        390
   CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
   GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
    6860       6870       6880       6890       6900       6910       6920

    400        410        420        430        440        450        460
   GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
   GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
    6930       6940       6950       6960       6970       6980       6990

          470        480        490        500        510        520
   GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
   AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
     7000       7010       7020       7030       7040       7050       7060

  530        540        550        560        570        580        590        600
   CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
   AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
   7070       7080       7090       7100       7110       7120       7130

          610        620        630        640        650        660
   AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
   AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
     7140       7150       7160       7170       7180       7190       7200

    670        680        690        700        710        720        730
   AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
   AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
     7210       7220       7230       7240       7250       7260       7270

         740        750        760        770        780        790
   AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA----CTGGGTAATCTT
   AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
   7280       7290       7300       7310       7320       7330       7340

  800        810        820        830        840        850        860
   GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
   GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
       7350       7360       7370       7380       7390       7400       7410

  870        880        890        900        910        920        930
   AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
```



```
   AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
        7420       7430       7440       7450       7460       7470

 940        950        960        970        980        X
   GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
   GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
 7480       7490       7500       7510       7520       7530       7540
```

7. US-08-162-407-5' (1-988)
   EBV Epstein-Barr virus (EBV) genome. The complete sequ

LOCUS EBV 172281 bp DNA circular VRL 04-JAN-1995
DEFINITION Epstein-Barr virus (EBV) genome. The complete sequence [1-10] was determined from DNA from B95-8 cells cloned by Arrand et al [11]. B95-8 is a productive marmoset lymphoblastoid cell line immortalized with human EBV from a mononucleosis patient.
ACCESSION V01555 J02070 K01729 K01730 V01554 X00498 X00499 X00784
KEYWORDS DNA polymerase; EBNA; genome; ribonucleotide reductase; tandem repeat; terminal repeat.
SOURCE Human herpesvirus 4.
  ORGANISM Human herpesvirus 4
Viridae; ds-DNA enveloped viruses; Herpesviridae; Gammaherpesviridae.
REFERENCE 1 (bases 1 to 172281)
  AUTHORS Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrell,B.G.
  TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome
  JOURNAL Nature 310, 207-211 (1984)
  STANDARD full automatic
REFERENCE 2 (bases 1 to 172281)
  AUTHORS Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.
  TITLE Sequence analysis and in vitro transcription of portions of the Epstein-Barr virus genome
  JOURNAL J. Cell. Biochem. 19, 267-274 (1982)
  STANDARD full automatic
REFERENCE 3 (bases 1 to 172281)
  AUTHORS Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.
  TITLE Homologous upstream sequences near Epstein-Barr virus promoters
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 80, 1565-1569 (1983)
  STANDARD full automatic
REFERENCE 4 (bases 1 to 172281)
  AUTHORS Farrell,P.J., Bankier,A.T., Seguin,C., Deininger,P.L. and Barrell,B.G.
  TITLE Latent and lytic cycle promoters of Epstein-Barr virus
  JOURNAL EMBO J. 2, 1331-1338 (1983)
  STANDARD full automatic
REFERENCE 5 (bases 142687 to 159853)
  AUTHORS Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.
  TITLE Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8 Epstein-Barr virus
  JOURNAL Mol. Biol. Med. 1, 21-45 (1983)
  STANDARD full automatic
REFERENCE 6 (bases 112620 to 125316)
  AUTHORS Seguin,C., Farrell,P.J. and Barrell,B.G.
  TITLE DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus
  JOURNAL Mol. Biol. Med. 1, 369-392 (1983)

STANDARD full automatic
REFERENCE 7 (bases 159853 to 172281)
AUTHORS Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J. and Barrell,B.G.
TITLE DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences
JOURNAL Mol. Biol. Med. 1, 425-445 (1983)
STANDARD full automatic
REFERENCE 8 (bases 87650 to 92703)
AUTHORS Biggin,M., Farrell,P.J. and Barrell,B.G.
TITLE Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus
JOURNAL EMBO J. 3, 1083-1090 (1984)
STANDARD full automatic
REFERENCE 9 (bases 76089 to 79808)
AUTHORS Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.
TITLE Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes
JOURNAL Nucleic Acids Res. 12, 5087-5099 (1984)
STANDARD full automatic
REFERENCE 10 (bases 1 to 172281)
AUTHORS Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeoch,D.
JOURNAL Unpublished
STANDARD full automatic
REFERENCE 11 (bases 1 to 172281)
AUTHORS Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and Griffin,B.E.
TITLE Molecular cloning of the complete Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments
JOURNAL Nucleic Acids Res. 9, 2999-3014 (1981)
STANDARD full automatic
REFERENCE 12 (bases 1 to 172281)
AUTHORS Kozak,M.
TITLE Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes
JOURNAL Nucleic Acids Res. 9, 5233-5262 (1981)
STANDARD full automatic
REFERENCE 13 (bases 7315 to 9312)
AUTHORS Yates,J., Warren,N., Reisman,D. and Sugden,B.
TITLE A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81, 3806-3810 (1984)
STANDARD full automatic
REFERENCE 14 (bases 45415 to 52824)
AUTHORS Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.
TITLE The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus
JOURNAL EMBO J. 3, 813-821 (1984)
STANDARD full automatic
REFERENCE 15 (bases 45644 to 52450)
AUTHORS Jeang,K.T. and Hayward,S.D.
TITLE Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
JOURNAL J. Virol. 48, 135-148 (1983)
STANDARD full automatic
REFERENCE 16 (bases 1 to 172281)

AUTHORS Farrell,P.J. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1984) to the EMBL/GenBank/DDBJ databases
STANDARD full automatic
REFERENCE 17 (bases 1 to 172281)
AUTHORS Bodescot,M. and Perricaudet,M.
TITLE Clustered alternative splice sites in Epstein-Barr virus RNAs
JOURNAL Nucleic Acids Res. 15, 5887 (1987)
STANDARD full automatic
REFERENCE 18 (bases 1 to 172281)
AUTHORS Laux,G., Perricaudet,M. and Farrell,P.J.
TITLE A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome
JOURNAL EMBO J. 7, 769-774 (1988)
STANDARD full automatic
REFERENCE 19 (bases 1 to 172281)
AUTHORS Farrell,P.J.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1988) to the EMBL/GenBank/DDBJ databases. Farrell P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG
STANDARD full automatic
COMMENT CDS
Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam H1 fragment in which they start. eg BALF3 is the third leftward frame starting in Bam H1 fragment A. BORF1 is the first rightward frame in Bam H1 fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.
SITEs of POLYA signals
This feature lists all occurences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog ATTAAA is only listed when it is found in a position close to the end of a major reading frame.
SITEs of DONOR and ACCEPT sequences
This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case.
Restriction enzyme SITEs.
Only the positions of the sites Bam HI (BAM) are listed.
RPT
This feature is used to define repetitive sequences.
SITE DEL
This feature defines deletions in B95-8 with respect to other

strains such as RAJI and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.
SITE HPN
Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurences noted.
ORGRPL
Denotes the region that encompasses an origin of replication (ori P).[13].
NUMBERING
The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has benn moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

NCBI gi: 59074

FEATURES Location/Qualifiers

source	1..172281
	/organism="Epstein-Barr virus"
	/strain="B95-8"
mRNA	58..272
	/note="exon 2 terminal protein RNA"
mRNA	360..458
	/note="exon 3 terminal protein RNA"
misc_feature	complement(535)
	/note="polyA signal: AATAAA"
mRNA	540..788
	/note="exon 4 terminal protein RNA"
mRNA	871..951
	/note="exon 5 terminal protein RNA"
mRNA	1026..1196
	/note="exon 6 terminal protein RNA"
promoter	complement(1192)
	/note="TATA: TATAAAT"
mRNA	1280..1495
	/note="exon 7 terminal protein RNA"
promoter	complement(1383)
	/note="TATA: CATAAAA"
mRNA	1574..1682
	/note="exon 8 terminal protein RNA"
promoter	1676
	/note="TATA: TATAAAG"
promoter	1691
	/note="TATA: TATTAAA BN-R1 late promoter before BNRF1, gives 4.1kb late RNA. Probably encodes non glycosylated 140kd protein in membrane antigen. Also two latent RNAs spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson et al, 1985). The longer one encodes terminal protein."
CDS	1736..5692
	/note="BNRF1 reading frame, 5 NXT/S; NCBI gi: 59075"
	/codon_start=1
	/translation="MEERGRETQMPVARYGGPFIMVRLFGQDGEANIQEERLYELLSD PRSALGLDPGPLIAENLLLVALRGTNNDPRPQRQERARELALVGILLGNGEQGEHLGT ESALEASGNNYVYAYGPDWMARPSTWSAEIQQFLRLLGATYVLRVEMGRQFGFEVHRS RPSFRQFQAINHLVLFDNALRKYDSGQVAAGFQRALLVAGPETADTRPDLRKLNEWVF GGRAAGGRQLADELKIVSALRDTYSGHLVLQPTETLDTWKVLSRDTRTAHSLEHGFIH AAGTIQANCPQLFMRRQHPGLFPFVNAIASSLGWYYQTATGPGADARAAARRQQAFQT RAAAECHAKSGVPVVAGFYRTINATLKGGEGLQPTMFNGELGAIKHQALDTVRYDYGH

YLIMLGPFQPWSGLTAPPCPYAESSWAQAAVQTALELFSALYPAPCISGYARPPGPSA
VIEHLGSLVPKGGLLLFLSHLPDDVKDGLGEMGPARATGPGMQQFVSSYFLNPACSNV
FITVRQRGEKINGRTVLQALGRACDMAGCQHYVLGSTVPLGGLNFVNDLASPVSTAEM
MDDFSPFFTVEFPPIQEEGASSPVPLDVDESMDISPSYELPWLSLESCLTSILSHPTV
GSKEHLVRHTDRVSGGRVAQQPGVGPLDLPLADYAFVAHSQVWTRPGGAPPLPYRTWD
RMTEKLLVSAKPGGENVKVSGTVITLGEQGYKVSLDLREGTRLAMAEALLNAACAPIL
DPEDVLLTLHLHLDPRRADNSAVMEAMTAASDYARGLGVKLTFGSASCPETGSSASNF
MTVVASVSAPGEFSGPLITPVLQKTGSLLIAVRCGDGKIQGGSLFEQLFSDVATTPRA
PEALSLKNLFRAVQQLVKSGIVLSGHDISDGGLVTCLVEMALAGQRGVTITMPVASDY
LPEMFAEHPGLVFEVEERSVGEVLQTLRSMNMYPAVLGRVGEQGPDQMFEVQHGPETV
LRQSLRLLLGTWSSFASEQYECLRPDRINRSMHVSDYGYNEALAVSPLTGKNLSPRRL
VTEPDPRCQVAVLCAPGTRGHESLLAAFTNAGCLCRRVFFREVRDNTFLDKYVGLAIG
GVHGARDSALAGRATVALINRFPALRDAILKFLNRPDTFSVALGELGVQVLAGLGAVG
STDNPPAPGVEVNVQRSPLILAPNASGMFESRWLNISIPATTSSVMLRGLRGCVLPCW
VQGSCLGLQFTNLGMPYVLQNAHQIACHFHSNGTDAWRFAMNYPRNPTEQGNIAGLCS
RDGRHLALLCDPSLCTDFWQWEHIPPAFGHPTGCSPWTLMFQAAHLWSLRHGRPSE"

misc_feature	complement(1795)
	/note="polyA signal: AATAAA"
misc_feature	3955
	/note="BAM: Bam H1 Nhet/h"
misc_feature	3994
	/note="BAM: Bam H1 h/C"
mRNA	5408..5856
	/note="exon 9 terminal protein RNA"
misc_feature	5841
	/note="polyA signal: AATAAA, end of 4.1kb late RNA and TP latent RNA."
misc_feature	5863
	/note="alternative end to TP cDNAs"
promoter	6097
	/note="TATA: TATAAGA"
misc_feature	6629..6795
	/note="Pol III RNA EBER 1"
promoter	complement(6823)
	/note="TATA: CATAAAT"
misc_feature	6956..7128
	/note="Pol III RNA EBER 2"
rep_origin	7315..9312
	/note="origin of replication, ori P (Yates et al, 1984, 1985)"
repeat_region	7421..8042
	/note="21x30bp repeats, binding sites for EBNA-1 (site I, Rawlins et al, 1985). Tandem repeat part of oriP (Reisman et al, 1985). Also functions as a cell type specific enhancer (Reisman et al, 1985; Lupton and Levine, 1985)"
promoter	7738
	/note="TATA: TATAAAT"
promoter	7888
	/note="TATA: TATAAAT"
promoter	8573
	/note="TATA: CATAAAT"
misc_feature	complement(8680)
	/note="polyA signal: AATAAA"
misc_feature	complement(8755)
	/note="polyA signal: AATAAA"
misc_feature	8962
	/note="polyA signal: AATAAA"
misc_feature	9021..9133
	/note="HPN: dyad symmetry, site II for EBNA-1 binding (Rawlins et al, 1985). Dyad symmetry part of oriP (Reisman

```
                et al, 1985)"
promoter        complement(9398)
                /note="TATA: TATAAAT"
promoter        9631
                /note="TATA: TATAAAT BC-R1 late promoter before BCRF1"
CDS             9675..10187
                /note="BCRF1 reading frame;  NCBI gi: 59076"
                /codon_start=1
                /translation="MERRLVVTLQCLVLLYLAPECGGTDQCDNFPQMLRDLRDAFSRV
                KTFFQTKDEVDNLLLKESLLEDFKGYLGCQALSEMIQFYLEEVMPQAENQDPEAKDHV
                NSLGENLKTLRLRLRRCHRFLPCENKSKAVEQIKNAFNKLQEKGIYKAMSEFDIFINY
                IEAYMTIKAR"
promoter        10076
                /note="TATA: GATAAAA"
misc_feature    complement(10148)
                /note="polyA signal: AATAAA"
misc_feature    10173
                /note="polyA signal: ATTAAA"
misc_feature    10257
                /note="polyA signal: AATAAA, end of 0.8kb late RNA from
                BCR1 and end of 1.6 kb late RNA, start unknown"
misc_feature    complement(10277)
                /note="polyA signal: AATAAA"
promoter        complement(10975)
                /note="TATA: CATAAAT"
promoter        11305
                /note="TATA: TACAAAA; BCR2 promoter for highly spliced
                EBNA latent RNAs."
mRNA            11336..11480
                /note="exon C1 of Bodescot et al (1986) RNAs"
promoter        11524
                /note="TATA: TATAATT"
misc_feature    complement(11587)
                /note="polyA signal: AATAAA"
promoter        complement(11606)
                /note="TATA: CATAAAT"
mRNA            11626..11657
                /note="exon C2 of Bodescot et al (1986) RNAs"
promoter        11796
                /note="TATA: TATAAGT"
promoter        complement(11799)
                /note="TATA: TATAAAA"
repeat_region   12001..15072
                /note="3072 repeat 1"
CDS             12541..13692
                /note="BCRF2 3072 repeat, reading frame 1;  NCBI gi:
                583873"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    13215
                /note="BAM: BamH1 C/W"
promoter        14352
                /note="*TATA: TATAAAG BWR1 one of the promoters for highly
                spliced EBNA and LP RNAs (Sample et al, 1986; Speck et al,
```



```
                1986)"
mRNA            14384..14410
                /note="*exon W0 of EBNA/LP RNAs"
mRNA            14554..14619
                /note="*exon W1 (also W66) part of leader protein (LP)
                gene. LP is also called EBNA-5 (Dillner et al, 1986) and
                EBNA-4 (Rowe et al, 1987)."
mRNA            14559..14619
                /note="*exon W1' (also W61) of EBNA/LP RNAs forms
                initiator met when fused to exon W0 or exon C2."
mRNA            14701..14832
                /note="*exon W2 (also W132) part of LP gene"
repeat_region   15073..18144
                /note="3072 repeat 2"
CDS             15613..16764
                /note="BWRF1 reading frame 2;  NCBI gi: 583874"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    16287
                /note="BAM: BamH1 W/W"
promoter        17424
                /note="TATA: TATAAAG"
mRNA            17626..17691
                /note="Exon W1"
mRNA            17773..17904
                /note="Exon W2"
repeat_region   18145..21216
                /note="3072 repeat 3"
CDS             18685..19836
                /note="BWRF1 reading frame 3;  NCBI gi: 583875"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    19359
                /note="BAM: BamH1 W/W"
promoter        20496
                /note="TATA: TATAAAG"
mRNA            20698..20763
                /note="Exon W1"
mRNA            20845..20976
                /note="Exon W2"
repeat_region   21217..24288
                /note="3072 repeat 4"
CDS             21757..22908
                /note="BWRF1 reading frame 4;  NCBI gi: 583876"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
```

```
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         22431
                     /note="BAM: BamH1 W/W"
promoter             23568
                     /note="TATA: TATAAAG"
mRNA                 23771..23835
                     /note="Exon W1"
mRNA                 23917..24048
                     /note="Exon W2"
repeat_region        24289..27360
                     /note="3072 repeat 5"
CDS                  24829..25980
                     /note="BWRF1 reading frame 5;  NCBI gi: 583877"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         25503
                     /note="BAM: BamH1 W/W"
promoter             26640
                     /note="TATA: TATAAAG"
mRNA                 26842..26907
                     /note="Exon W1"
mRNA                 26989..27120
                     /note="Exon W2"
repeat_region        27361..30432
                     /note="3072 repeat 6"
CDS                  27901..29052
                     /note="BWRF1 reading frame 6;  NCBI gi: 583878"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         28575
                     /note="BAM: BamH1 W/W"
promoter             29712
                     /note="TATA: TATAAAG"
mRNA                 29914..29979
                     /note="Exon W1"
mRNA                 30061..30192
                     /note="Exon W2"
repeat_region        30433..33504
                     /note="3072 repeat 7"
CDS                  30973..32124
                     /note="BWRF1 reading frame 7;  NCBI gi: 583879"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
```



```
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         31647
                     /note="BAM: BamH1 W/W"
promoter             32784
                     /note="TATA: TATAAAG"
mRNA                 32986..33051
                     /note="Exon W1"
mRNA                 33133..33264
                     /note="Exon W2"
repeat_region        33505..36576
                     /note="3072 repeat 8"
CDS                  34045..35196
                     /note="BWRF1 reading frame 8;  NCBI gi: 583880"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         34719
                     /note="BAM: BamH1 W/W"
promoter             35856
                     /note="TATA: TATAAAG"
mRNA                 36058..36123
                     /note="Exon W1"
mRNA                 36205..36336
                     /note="Exon W2"
repeat_region        36577..39648
                     /note="3072 repeat 9"
CDS                  37117..38268
                     /note="BWRF1 reading frame 9;  NCBI gi: 583881"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         37791
                     /note="BAM: BamH1 W/W"
promoter             38928
                     /note="TATA: TATAAAG"
mRNA                 39130..39195
                     /note="Exon W1"
mRNA                 39277..39408
                     /note="Exon W2"
repeat_region        39649..42720
                     /note="3072 repeat 10"
CDS                  40189..41340
                     /note="BWRF1 reading frame 10;  NCBI gi: 583882"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
```

```
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    40863
                /note="BAM: BamH1 W/W"
promoter        42000
                /note="TATA: TATAAAG"
mRNA            42202..42267
                /note="Exon W1"
mRNA            42349..42480
                /note="Exon W2"
repeat_region   42721..45792
                /note="3072 repeat 11"
CDS             43261..44412
                /note="BWRF1 reading frame 11;  NCBI gi: 583883"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    43935
                /note="BAM: BamH1 W/W"
promoter        45072
                /note="TATA: TATAAAG"
mRNA            45274..45339
                /note="Exon W1"
misc_feature    45415..52824
                /note="DEL: DAUDI deletion (Jones et al, 1984)"
mRNA            45421..45552
                /note="Exon W2"
misc_feature    45644..52450
                /note="DEL: P3HR1 deletion (Jeang and Hayward, 1983)"
repeat_region   45793..47643
                /note="3072 repeat 12"
CDS             46333..47484
                /note="BWRF1 reading frame 12;  NCBI gi: 583884"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    47007
                /note="BAM: BamH1 W/Y"
mRNA            47761..47793
                /note="Exon Y1 Bodescot et al, 1984"
promoter        47831
                /note="TATA: TATAAGT"
mRNA            47878..47999
                /note="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and
                Strominger,1985), last common exon"
misc_feature    complement(48023)
                /note="polyA signal: AATAAA"
CDS             48386..50032
```



```
                /note="Coding exon for EBNA-2 (Sample et al,1986)"
                /codon_start=1
mRNA            48386..48444
                /note="exon Bodescot et al, 1984"
CDS             48429..49967
                /note="BYRF1, encodes EBNA-2 (Dambaugh et al, 1984;
                Dillner et al, 1984);  NCBI gi: 583885"
                /codon_start=1
                /translation="XCGWAGTCQQPSKHPRLCFALSAAIMPTFYLALHGGQTYHLIVD
                TDSLGNPSLSVIPSNPYQEQLSDTPLIPLTIFVGENTGVPPPLPPPPPPPPPPPPPP
                PPPPPPPPPPSPPPPPPPPPPQRRDAWTQEPSPLDRDPLGYDVGHGPLASAMRMLW
                MANYIVRQSRGDRGLILPQGPQTAPQARLVQPHVPPLRPTAPTILSPLSQPRLTPPQP
                LMMPPRPTPPTPLPPATLTVPPRPTRPTTLPPTPLLTVLQRPTELQPTPSPPRMHLPV
                LHVPDQSMHPLTHQSTPNDPDSPEPRSPTVFYNIPPMPLPPSQLPPPAAPAQPPPGVI
                NDQQLHHLPSGPPWWPPICDPPQPSKTQGQSRGQSRGRGRGRGRGRGKGKSRDKQRKP
                GGPWRPEPNTSSPSMPELSPVLGLHQGQGAGDSPTPGPSNAAPVCRNSHTATPNVSPI
                HEPESHNSPEAPILFPDDWYPPSIDPADLDESWDYIFETTESPSSDEDYVEGPSKRPR
                PSIQ"
repeat_region   48678..48800
                /note="14 x CCCCCACCA repeats"
misc_feature    48848
                /note="BAM: BamH1 Y/H"
promoter        49350
                /note="TATA: TATAACA"
promoter        complement(49353)
                /note="TATA: TATAAAA"
repeat_region   49525..49578
                /note="9 x GGGGCA repeats"
mRNA            49852..50032
                /note="exon (Bodescot et al 1984)"
misc_feature    50003
                /note="polyA signal: AATAAA, end of Bodescot T1 RNA and
                EBNA-2 RNA (3.0kb latent RNA in IB4 cells)"
promoter        complement(50156)
                /note="TATA: TATAAGT"
misc_feature    complement(50317)
                /note="polyA signal: AATAAA, end of 2.5kb early RNA from
                52817"
repeat_region   50578..52115
                /note="12 x 125bp repeats"
misc_feature    complement(50578..52557)
                /note="BHLF1 early reading frame"
misc_feature    52654..53697
                /note="region homologous to Eco R1 C of Raji"
promoter        complement(52817)
                /note="TATA: GATAAAA promoter for 2.5kb early RNA
                containing BHLF1 (Jeang and Hayward, 1983; Freese et
                al,1983)"
promoter        53759
                /note="TATA: TATTAAC likely promoter for class III and IV
                early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature    53895
                /note="DONOR: CGGGTAACT donor for splice to 54335 in class
                IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature    54335
                /note="ACCEPT: TTTTCTAG acceptor from 48444 in class I,
                47999 in class II, and 53895 in class IV early RNAs
                encoding BHRF1 (Pearson et al, 1987)"
misc_feature    54376..54948
                /note="BHRF1 reading frame, limited homolgy to bcl-2 gene.
```

```
                Early gene in B95-8 cells and part of restricted EA
                complex."
promoter        54591
                /note="TATA: TATAACA"
promoter        complement(54594)
                /note="TATA: TATAAAT"
misc_feature    54853
                /note="BAM: BamH1 H/F"
misc_feature    complement(54929)
                /note="polyA signal: AATAAA"
promoter        complement(54977)
                /note="TATA: TATAAAG"
misc_feature    55518
                /note="polyA signal: AATAAA, 3' end of 2.5kb, 1.9kb, 1.7kb
                and 0.6kb early RNAs"
misc_feature    complement(55982..56935)
                /note="BFLF2 reading frame, 4 NXT/S, homologous to RF 27
                in VZV and HFRF2 in CMV"
misc_feature    complement(55990)
                /note="polyA signal: AATAAA, 3' end of 2.3kb and 1.1kb
                early RNAs from 58568 and 57081"
promoter        complement(56132)
                /note="TATA: TATAAAG"
CDS             complement(56948..58525)
                /note="BFLF1 reading frame, 2 NXT/S homologous to RF 26 in
                VZV and HFRF1 in CMV;  NCBI gi: 59090"
                /codon_start=1
                /translation="MAHKVTSANEPNPLTGKRLSSCPLTRSGVTEVAQIAGRTPKMED
                FVPWTVDNLKSQFEAVGLLMAHSYLPANAEEGIAYPPLVHTYESLSPASTCRVCDLLD
                TLVNHSDAPVAFFEDYALLCYYCLNAPRAWISSLITGMDFLHILIKYFPMAGGLDSLF
                MPSRILAIDIQLHFYICRCFLPVSSSDMIRNANLGYYKLEFLKSILTGQSPANFCFKS
                MWPRTTPTFLTLPGPRTCKDSQDVPGDVGRGLYTALCCHLPTRNRVQHPFLRAEKGGL
                SPEITTKADYCGLLLGTWQGTDLLGGPGHHAIGLNAEYSGDELAELALAITRPEAGDH
                SQGPCLLAPMFGLRHKNASRTICPLCESLGAHPDAKDTLDRFKSLILDSFGNNIKILD
                RIVFLIKTQNTLLDVPCPRLRAWLQMCTPQDFHKHLFCDPLCAINHSITNPSVLFGQI
                YPPSFQAFKAALAAGQNLEQGVCDSLITLVYIFKSTQVARVGKTILVDVTKELDVVLR
                IHGLDLVQSYQTSQVYV"
promoter        complement(57081)
                /note="TATA: TATTTAA before BFLF2; BFL2 promoter gives
                1.1kb early RNA"
promoter        complement(58088)
                /note="TATA: GATAAAA"
promoter        complement(58568)
                /note="TATA: TATTAAA before BFLF1, BFL1 promoter gives
                2.3kb early RNA"
promoter        58832
                /note="TATA: TATAAAA before BFRF1"
CDS             58891..59901
                /note="BFRF1 early reading frame, 1 NXT/S, homologous to
                HFLF4 in CMV;  NCBI gi: 59091"
                /codon_start=1
                /translation="MASPEERLLDELNNVIVSFLCDSGSLEVERCSGAHVFSRGSSQP
                LCTVKLRHGQIYHLEFVYKFLAFKLKNCNYPSSPVFVISNNGLATTLRCFLHEPSGLR
                SGQSGPCLGLSTDVDLPKNSIIMLGQDDFIKFKSPLVFPAELDLLKSMVVCRAYITEH
                RTTMQFLVFQAANAQKASRVMDMISDMSQQLSRSGQVEDTGARVTGGGGPRPGVTHSG
                CLGDSHVRGRGGWDLDNFSEAETEDEASYAPWRDKDSWSESEAAPWKKELVRHPIRRH
                RTRETRRMRGSHSRVEHVPPETRETVVGGAWRYSWRATPYLARVLAVTAVALLLMFLR
                WT"
CDS             59610..61583
                /note="BFRF2 early reading frame, homologous to HFLF5 in
```



```
                CMV;  NCBI gi: 583886"
                /codon_start=1
                /translation="XRRRGELRSLEGQRLVVGIRGGAVEEGTREAPHPQAPDTRDSPY
                AREPFTGGTRAPRDPGDGGGGSMALFLARHTLSGTGAGCHGRGPAPDVSEVDLTLQAL
                GERGFSRLLDLGLACLDLSYVEMREFVVWGRPPASEAAVASTPGSLFRSHSSAYWLSE
                VERPGGLVRWARSQTSPSSLTLAPHLGPSLLSLSVVTGGGCGAVAFCNAFFLAYFLVV
                RSVFPAFSDRIAAWICDRSPFCENTRAVARGYRGLVKRFLAFVFERSSYDPPLLRQNS
                RPVERCFAIKNYVPGLDSQSCVTVPSFSRWAQSHASELDPREIRDRVTPATAPSFVAD
                HASALLASLQKKASDTPCGNPIQWMWYRLLVNSCLRSAHCLLPIPAVSEGGRKTGGGV
                GEELVGAGGPCLSRDVFVAIVSRNVLSCLLNVPAAGPRAYKCFRSHASRPVSGPDYPP
                LAVFCMDCGYCLNFGKQTGVGGRLNSFRPTLQFYPRDQKEKHVLTCHASGRVYCSNCG
                SAAVGCQRLAEPPSARSGWRPRIRAVLPHNAAYELDRGSRLLDAIIPCLGPDRTCMRP
                VVLRGVTVRQLLYLTLRTEARAVCSICQQRQAPEDARDEPHLFSSCLEVELPPGERCA
                GCRLYQTRYGTPAAQAHPPGEAGGGFSRQSPAS"
promoter        complement(61062)
                /note="TATA: GATAAAA"
promoter        61344
                /note="TATA: TATTTAA before BFRF3"
CDS             61456..62037
                /note="BFRF3 early reading frame;  NCBI gi: 583887"
                /codon_start=1
                /translation="XCHLVSGVRAAVSIRRVMARRLPKPTLQGRLEADFPDSPLLPKF
                QELNQNNLPNDVFREAQRSYLVFLTSQFCYEEYVQRTFGVPRRQRAIDKRQRASVAGA
                GAHAHLGGSSATPVQQAQAAASAGTGALASSAPSTAVAQSATPSVSSSISSLRAATSG
                ATAAASAAAAVDTGSGGGGQPHDTAPRGARKKQ"
misc_feature    complement(62068)
                /note="polyA signal: AATAAA"
misc_feature    62069
                /note="polyA signal: AATAAA, 3' end of 10, 6.5, 3.7, 3.4,
                3.1, 2.5 and 0.8kb early RNAs"
CDS             complement(62078..71527)
                /note="BPLF1 reading frame, 1 NXT/S, analogous to VZV
                RF22;  NCBI gi: 59094"
                /codon_start=1
                /translation="MSNGDWGQSQRTRGTGPVRGIRTMDVNAPGGGSGGSALRILGTA
                SCNQAHCKFGRFAGIQCVSNCVLYLVKSFLAGRPLTSRPELDEVLDEGARLDALMRQS
                GILKGHEMAQLTDVPSSVVLRGGGRVHIYRSAEIFGLVLFPAQIANSAVVQSLAEVLH
                GSYNGVAQFILYICDIYAGAIIIETDGSFYLFDPHCQKDAAPGTPAHVRVSTYAHDIL
                QYVGAPGAQYTCVHLYFLPEAFETEDPRIFMLEHYGVYDFYEANGSGFDLVGPELVSS
                DGEAAGTPGADSSPPVMLPFERRIIPYNLRPLPSRSFTSDSFPAARYSPAKTNSPPSS
                PASAAPASAAPASAAPASAAPASAAPASAAPASAAPASAAPASSPPLFIPIPGLGHTP
                GVPAPSTPPRASSGAAPQTPKRKKGLGKDSPHKKPTSGRRLPLSSTTDTEDDQLPRTH
                VPPHRPPSAARLPPPVIPIPHQSPPASPTPHPAPVSTIAPSVTPSPRLPLQIPIPLPQ
                AAPSNPKIPLTTPSPSPTAAAAPTTTTLSPPPTQQQPPQSAAPAPSPLLPQQQPTPSA
                APAPSPLLPQQQPPPSAARAPSPLPPQQQPLPSATPAPPPAQQLPPSATTLEPEKNHP
                PAADRAGTEISPSPPFGQQPSFGDDASGGSGLVRYLSDLEEPFLSMSDSEEAESDLAS
                DIPTTEDEDMFEDEVFSNSLESGSSAPTSPITLDTARSQYYQTTFDIETPEMDFVPLE
                SNIARIAGHTYQEQAIVYDPASNREVPEADALSMIDYLLVTVVLEQGLIRSRDRSSVL
                NLLEFLKDWSGHLQVPTLDLEQLLTSELNIQNLANMLSENKGRAGEFHKHLAAKLEAC
                LPSLATKDAVRVDAGAKMLAEIPQLAESDDGKFDLEAARRRLTDLLSGGDQEAGEGGG
                EPEDNSIYRGPHVDVPLVLDDESWKRLLSLAEAARTAVARQQAGVDEEDVRFLALLTA
                IEYGAPPAASVPPFVHNVAVRSKNAALHVRRCTADIRDKVASAASDYLSYLEDPSLPT
                VMDFDDLLTHLRHTCQIIASLPLLNIRYTSIEWDYRELLYLGTALSDMSGIPWPLERV
                EEDDPSIAPLPEFETVAKKQKELETTRENEKRLRTILDDIEAMLGLAGVASAPGAPIS
                PASPSATPANHDNPEATPPLADTAALTIPVIEKYIANAGSIVGAAKNPTYIRLRDTIQ
                QIVRSKKYLMNILKSITFYTIDNYIASFEESIDHLYRDLPVLDPEVQDGIDRILDPMV
                SEALHTFEMGNRLTLEPARLVALQNFATHSTLKETAAAVNLLPGLLAVYDATITGQAP
                EDALRLLSGLQNQLSQTLIPGKLKKRFLSYLQKLKNNNNDQLRQKEVQAWRLEAEGFK
                PATEEQLEAFLDTAPNKELKRQYEKKLRQLMETGRKEKEKLREQEDKERQERRAREAN
                EAWARIRKALGARPEPAPTSPDDWNTLLASLLPDNTDSAAAAAAAVARNTDILDSLTQ
```

```
                     ILAAMLLGITRVRRERLRSLLVDDGGAAERMEAAEPGWFTDIETGPLARLDAWPATPA
                     ATAKEGGGGRGAEEAAGALFRARTAADAIRSALAQTRQALQSPDMKSAVVNTDLEAPY
                     AEYERGLAGLLEKRRAAEAALTAIVSEYVDRTLPEATNDPGQANLPPPPTIPQATAPP
                     RLASDSALWPKKPQLLTRRERDDLLQATGDFFSELLTEAEAAEVRALEEQVRESQTLM
                     AKAHEMAASTRRGFHTALEAVLSRSRDEAPDDELRSLLPSPPKAPVQAPLEAALARAA
                     AGNGSWPYRKSLAAAKWIRGICEAVRGLSEGALALAGGAGAWLNLAAAADGEIHELTR
                     LLEVEGMAQNSMDGMEELRLALATLDPKRVAGGKETVADWKRRLSRLEAIIQEAQEES
                     QLQGTLQDLVTQARGHTDPRQLKIVVEAARGLALGASAGSQYALLKDKLLRYASAKQS
                     FLAFYETAQPTVFVKHPLTNNLPLLITISAPPTGWGNGAPTRRAQFLAAAGPAKYAGT
                     LWLETESPCDPLNPAYVSADTQEPLNYIPVYHNFLEYVMPTVLENPEAFSLTPAGRPQ
                     AIGPPQDDQERRRRTLASVASARLSAAAADSYWDTWPDVESNAGELLREYVSAPKALM
                     EDLADNPIVAMTLLAHASLIASRNHPPYPAPATDREVILLEQREMMALLVGTHPAYAA
                     AFLGAPSFYAGLGLVSALARDGGLGDLLSDSVLTYRLVRSPASGRGGMPSTTRGSNDG
                     EDARRLTRHRIAGPPTGFIFFQDAWEEMDTRAALWPHPEFLGLVHNQSTARARACMLL
                     LARRCFAPEALQQLWHSLRPLEGPVAFQDYLRDFVKQAYTRGEELPRAEGLEVPRETP
                     SSYGTVTGRALRNLMPYGTPITGPKRGSGDTIPVSVFEAAVAAAFLGRPLTLFVSSQY
                     LFNLKTLGQVRVVAPLLYCDGHSEPFRSLVETISLNFLQDLDGYSESFEPEMSIFARQ
                     AVWLRELLTEARAAKPKEARPPTVAILANRKNIIWKCFTYRHNLPDVQFYFNAAGASR
                     WPTDVLNPSFYEHEDPPLPVGYQLPPNPRNVQELFSGFPPRVGHGLVSGDGFQSADNT
                     PASSDRLQQLGGGETDQGEKGSTTAESEASGPPSPQSPLLEKVAPGRPRDWLSPTSSP
                     RDVTVTPGLAAPITLPGPRLMARPYFGAETRASESPDRSPGSSPRPWPKDSLELLPQP
                     APQQPPSSPWASEQGPIVYTLSPHSTPSTASGSQKKHTIQIPGLVPSQKPSYPPSAPY
                     KPGQSTGGIAPTPSAASLTTFGLQPQDTQASSQDPPYGHSIMQREKKQQGGREEAAEI
                     RPSATRLPTAVGLRPRAPVVAAGAAASATPAFDPGEAPSGFPIPQAPALGSGLAAPAH
                     TPVGALAPRPQKTQAQRPQDAAALPTPTIKAVGARPVPKATGALAAGARPRGQPTAAP
                     PSAASPPRVSLPVRSRQQQSPAIPLPPMHSGSEPGARPEVRLSQYRHAGPQTYTVRKE
                     APPSAASQLPKMPKCKDSMYYPPSGSARYPAPFQALSFSQSVASPAPSSDQTTLLWNT
                     PSVVTQFLSIEDIIREVVTGGSTSGDLVVPSGSPSSLSTAAPEQDLRYSLTLSQASRV
                     LSRFVSQLRRKLERSTHRLIADLERLKFLYL"
misc_feature         62249
                     /note="BAM: BamH1 F/Q"
misc_feature         62430..62477
                     /note="Site III for EBNA-1 binding (Rawlins et al, 1985)"
misc_feature         66121
                     /note="BAM: BamH1 Q/U"
mRNA                 67477..67649
                     /note="Exon in EBNA-1 RNA (Speck and Strominger,1985) and
                     cDNA clone T4 (Bodescot et al, 1986)"
misc_feature         69410
                     /note="BAM: BamH1 U/P"
repeat_region        69684..69930
                     /note="5 x 51bp repeats"
repeat_region        70387..70521
                     /note="9 x 15bp repeat"
promoter             70750
                     /note="TATA: CATAAAA"
CDS                  complement(71520..75239)
                     /note="BOLF1 reading frame, 1 NXT/S analogous to VZV RF
                     21;  NCBI gi: 59096"
                     /codon_start=1
                     /translation="MASAMESDSSGGSGGADAQPPLAEVDGGLARVTRQLLLSGDDPA
                     ARLRALMPLELGIFGLGDLAQPVLVRDFLNTLTLMSGHAYPAAVLRHHAYYLLRAASF
                     SRRSFGLGHLEAALDVLASSLPPTTASPATDDPLDGSRLIAETRALAAAYRRIIEEGS
                     GEVLAVSGPTATFAFVEELVADTYLARWDAFPREGLSFYAFNAAKTTLGRWLVTVYAE
                     TNRYPWAAAGQGQPTAADIKAMAVELVEHSGGGAGGGEGEESGGGLFHRPESLSSVVA
                     SLPLARRAVEILGVYAEASGGQTPPVAAVPVLAFDAARLRLLEPSGALFYDYVYEAL
                     LWDQTYGVPDSVIEAFLAGMAAEMEALAARVQEAAGSRASFSPAAIEQVATVLLSAGL
                     NETVAGDYAMMLASVPRVSRSRWRWLEATAALLESLSGFALHFFRLLPTASPTSRFAR
                     VARAAYLRAEAEAVDRRARRTSGPSTPAAAPAATAVGVGAAADPWDAVTPLRIFIVPP
                     PAAEYEQVAGDLSSELLRSLLWVRYSRLWQAPAPAPALPCKPPLLPGEQGRRQWTAAV
                     AAAPRTDVEAYCRSLRAGQTARADPAYVHSPFFPAAFIEFQIWPALRRVLSNELPKTR
                     SLAALRWLVSFGSDLALPSPELTRARRPLELIYATVWEIYDGAPPMPGESPQAVGLRP
                     LNLEGEGKAGDAGAEGAEDEEGGGPWGLSSHDAVLRIMDAVREVSGIISETISASERA
                     AEAPPLAWPTSLFSLLFTLRYSTTAESLGLATRRFLVSGETLSEDISRLTGAAWRLCS
                     RPLLYDAETGRVQIPLATEEEEEAVVAVKEKSVSSSPRHYSTDLQTLKSVVEGIQDVC
                     RDAAARWALATADTATLRRRLLVPALRESRGIADHPLWAHTSEPLRPDLEELNERVEH
                     ALELGYSLTGALRRSVAYRFRDYTFARLFQPPAIDAERAEAIVRRDARPPPVFIPAPR
                     RLPQGGADTPPPLSMDDILYLGKSICKALVDVLDHHPAAPETTPIKTYTPAMDLNPEQ
                     ITVTPRSPSVLAAFARTARVQTHHLVPALTDDSPSPVGQTPPPFRILPAKKLAAILLG
                     NGRNASKRRASRDLSPPPHGRWRAVLDSSPFSFSSSDFSDQDEGEGGEADLRGVPGGG
                     GEGAYEEDRERPSDIDTAARAQKVETSCPRRRSPRTTPSPSRRASGGGGPDRGEAEAH
                     TYPPYLSAAAAASRVRPRTRRGATRRPPRPTAEDE"
promoter             complement(72192)
                     /note="TATA: TATTAAA before BPLF1"
misc_feature         73468
                     /note="BAM: BamH1 P/O"
promoter             75017
                     /note="TATA: TATTTAA BO-R1 late promoter before BORF1,
                     gives 3.9kb late RNA"
CDS                  75238..76332
                     /note="BORF1 late reading frame, 2 NXT/S homologous to VZV
                     RF20;  NCBI gi: 59095"
                     /codon_start=1
                     /translation="MKVQGSVDRRRLQRRIAGLLPPPARRLNISRGSEFTRDVRGLVE
                     EHAQASSLSAAAVWRAGLLAPGEVAVAGGGSGGGSFSWSGWRPPVFGDFLIHASSFNN
                     AEATGTPLFQFKQSDPFSGVDAVFTPLSLFILMNHGRGVAARVEAGGGLTRMANLLYD
                     SPATLADLVPDFGRLVADRRFHNFITPVGPLVENIKSTYLNKITTVVHGPVVSKAIPR
                     STVKVTVPQEAFVDLDAWLSGGAGGGGGVCFVGGLGLQPCPADARLYVALTYEEAGPR
                     FTFFQSSRGHCQIMNILRIYYSPSIMHRYAVVQPLHIEELTFGAVACLGTFSATDGWR
                     RSAFNYRGSSLPVVEIDSFYSNVSDWEVIL"
promoter             complement(75322)
                     /note="TATA: TATTTAG before BOLF1"
promoter             75819
                     /note="TATA: TATAAAG"
misc_feature         75838
                     /note="polyA signal: AATAAA"
misc_feature         complement(76126)
                     /note="polyA signal: AATAAA"
promoter             76169
                     /note="TATA: TACATAT BO-R2 early promoter before BORF2,
                     gives 2.8kb RNA"
misc_feature         complement(76300)
                     /note="polyA signal: AATAAA"
CDS                  76407..78887
                     /note="BORF2 early reading frame, 2 NXT/S. Homology HSV
                     140K ribonucleotide reductase (Gibson et al, 1984) and RF
                     19 VZV;  NCBI gi: 59097"
                     /codon_start=1
                     /translation="MATTSHVEHELLSKLIDELKVKANSDPEADVLAGRLLHRLKAES
                     VTHTVAEYLEVFSDKFYDEEFFQMHRDELETRVSAFAQSPAYERIVSSGYLSALRYYD
                     TYLYVGRSGKQESVQHFYMRLAGFCASTTCLYAGLRAALQRARPEIESDMEVFDYYFE
                     HLTSQTVCCSTPFMRFAGVENSTLASCILTTPDLSSEWDVTQALYRHLGRYLFQRAGV
                     GVGVTGAGQDGKHISLLMRMINSHVEYHNYGCKRPVSVAAYMEPWHSQIFKFLETKLP
                     ENHERCPGIFTGLFVPELFFKLFRDTPWSDWYLFDPKDAGDLERLYGEEFEREYYRLV
                     TAGKFCGRVSIKSLMFSIVNCAVKAGSPFILLKEACNAHFWRDLQGEAMNAANLCAEV
                     LQPSRKSVATCNLANICLPRCLVNAPLAVRAQRADTQGDELLLALPRLSVTLPGEGAV
                     GDGFSLARLRDATQCATFVVACSILQGSPTYDSRDMASMGLGVQGLADVFADLGWQYT
                     DPPSRSLNKEIFEHMYFTALCTSSLIGLHTRKIFPGFKQSKYAGGWFHWHDWAGTDLS
                     IPREIWSRLSERIVRDGLFNSQFIALMPTSGCAQVTGCSDAFYPFYANASTKVTNKEE
                     ALRPNRSFWRHVRLDDREALNLVGGRVSCLPEALRQRYLRFQTAFDYNQEDLIQMSRD
```

```
                RAPFVDQSQSHSLFLREEDAARASTLANLLVRSYELGLKTIMYYCRIEKAADLGVMEC
                KASAALSVPREEQNERSPAEQMPPRPMEPAQVAGPVDIMSKGPGEGPGGWCVPGGLEV
                CYKYRQLFSEDDLLETDGFTERACESCQ"
misc_feature    77835
                /note="BAM: Bam H1 O/a"
promoter        78804
                /note="TATA: TATAAGT Ba-R1 early promoter before BaRF1,
                gives 3.5kb RNA"
misc_feature    78883
                /note="polyA signal: AATAAA, end of 3.9kb late RNA from
                75017 and 2.8kb early RNA from 76169"
misc_feature    complement(78896)
                /note="polyA signal: AATAAA"
CDS             78900..79808
                /note="BaRF1 early reading frame, homologous to HSV 38K
                ribonucleotide reductase (Gibson et al, 1984) and RF 18
                VZV;  NCBI gi: 59098"
                /codon_start=1
                /translation="MSKLLYVRDHEGFACLTVETHRNRWFAAHIVLTKDCGCLKLLNE
                RDLEFYKFLFTFLAMAEKLVNFNIDELVTSFESHDIDHYYTEQKAMENVHGETYANIL
                NMLFDGDRAAMNAYAEAIMADEALQAKISWLRDKVAAAVTLPEKILVFLLIEGIFFIS
                SFYSIALLRVRGLMPGICLANNYISRDELLHTRAASLLYNSMTAKADRPRATWIQELF
                RTAVEVETAFIEARGEGVTLVDVRAIKQFLEATADRILGDIGQAPLYGTPPPKDCPLT
                YMTSIKQTNFFEQESSDYTMLVVDDL"
promoter        complement(79495)
                /note="TATA: TATAACA"
misc_feature    79537
                /note="BAM: Bam H1 a/M"
promoter        79840
                /note="TATA: CATAAAT BM-R1 early promoter before BMRF1,
                gives 2.5kb RNA"
CDS             79899..81113
                /note="BMRF1 early reading frame. Early antigen protein
                recognised by R3 monoclonal (Pearson et al 1983; Cho et
                al, 1985a);  NCBI gi: 59099"
                /codon_start=1
                /translation="METTQTLRFKTKALAVLSKCYDHAQTHLKGGVLQVNLLSVNYGG
                PRLAAVANAGTAGLISFEVSPDAVAEWQNHQSPEEAPAAVSFRNLAYGRTCVLGKELF
                GSAVEQASLQFYKRPQGGSRPEFVKLTMEYDDKVSKSHHTCALMPYMPPASDRLRNEQ
                MIGQVLLMPKTASSLQKWARQQGSGGVKVTLNPDLYVTTYTSGEACLTLDYKPLSVGP
                YEAFTGPVAKAQDVGAVEAHVVCSVAADSLAAALSLCRIPAVSVPILRFYRSGIIAVV
                AGLLTSAGDLPLDLSVILFNHASEEAAASTASEPEDKSPRVQPLGTGLQQRPRHTVSP
                SPSPPPPPRTPTWESPARPETPSPAIPSHSSNTALERPLAVQLARKRTSSEARQKQKH
                PKKVKQAFNPLI"
promoter        80779
                /note="TATA: TATTTAA BM-R2 late promoter before BMRF2"
misc_feature    complement(80782)
                /note="polyA signal: AATAAA"
promoter        80832
                /note="TATA: GATAAAA, possible promoter for 1.4kb late RNA
                encoding BMRF2"
CDS             81118..82191
                /note="BMRF2 early reading frame;  NCBI gi: 59100"
                /codon_start=1
                /translation="MFSCKQHLSLGACVFCLGLLASTPFIWCFVFANLLSLEIFSPWQ
                THVYRLGFPTACLMAVLWTLVPAKHAVRAVTPAIMLNIASALIFFSLRVYSTSTWVSA
                PCLFLANLPLLCLWPRLAIEIVYICPAIHQRFFELGLLLACTIFALSVVSRALEVSAV
                FMSPFFIFLALGSGSLAGARRNQIYTSGLERRRSIFCARGDHSVASLKETLHKCPWDL
                LAISALTVLVVCVMIVLHVHAEVFFGLSRYLPLFLCGAMASGGLYLGHSSIIACVMAT
                LCTLTSVVVYFLHETLGPLGKTVLFISIFVYYFSGVAALSAAMRYKLKKFVNGPLVHL
```



```
                RVVYMCCFVFTFCEYLLVTFIKS"
promoter        81751
                /note="TATA: CATAAAT"
misc_feature    82180
                /note="polyA signal: ATTAAA, end of 3.5kb early RNA from
                78804, 2.5kb early RNA from 79840 and 1.4kb late RNA"
promoter        complement(82311)
                /note="TATA: CATAAAT"
repeat_region   82319..82461
                /note="2x71bp repeats"
CDS             complement(82743..84122)
                /note="BMLF1 early reading frame. Diffuse early antigen
                (Cho et al, 1985b). Also homologous to RF 4 VZV and IE63
                of HSV.(BSLF2 + BMLF1) is also called EB2
                (Chevallier-Greco et al, 1986). General transactivator of
                transcription (Lieberman et al, 1986).  NCBI gi: 583888"
                /codon_start=1
                /translation="XSHILELEAVSDTNTDCDLDPMEGSEEHSTDGEISSSEEEDEDP
                TPAHAIPARPSSVVITPTSASFVIPRKKWDLQDKTVTLHRSPLCRDEDKEETGNSSY
                TRGHKRRRGEVHGCTDESYGKRRHLPPGARAPRAPRAPRVPRAPRSPRAPRSNRATRG
                PRSESRGAGRSTRKQARQERSQRPLPNKPWFDMSLVKPVSKITFVTLPSPLASLTLEP
                IQDPFLQSMLAVAAHPEIGAWQKVQPRHELRRSYKTLREFFTKSTNKDTWLDARMQAI
                QNAGLCTLVAMLEETIFWLQEITYHGDLPLAPAEDILLACAMSLSKVILTKLKELAPC
                FLPNTRDYNFVKQLFYITCATARQNKVVETLSSSYVKQPLCLLAAYAAVAPAYINANC
                RRRHDEVEFLGHYIKNYNPGTLSSLLTEAVETHTRDCRSASCSRLVRAILSPGTGSLG
                LFFVPGLNQ"
misc_feature    complement(82747)
                /note="polyA signal: AATAAA"
repeat_region   83640..83729
                /note="10x9bp repeats"
misc_feature    complement(84122)
                /note="ACCEPT: CTCCCCTCTGCAG acceptor in spliced form of
                BMLF1 RNA"
misc_feature    complement(84227)
                /note="DONOR: CAGGTAAGA donor in spliced form of BMLF1
                RNA"
CDS             complement(84229..84288)
                /note="BSLF2 early reading frame in 5' exon of spliced RNA
                encoding BMLF1;  NCBI gi: 59102"
                /codon_start=1
                /translation="MVPSQRLSRTSSISSNEDPA"
misc_feature    84233
                /note="BAM: Bam H1 M/S"
CDS             complement(84257..86881)
                /note="BSLF1 reading frame, homologous to RF 6 VZV;  NCBI
                gi: 59103"
                /codon_start=1
                /translation="MSAPVVIKALVASNTDIAEAILDAILSRPDEGFRLFCLCHNASP
                LHHVAGSLVELQLHLPKKRLTSQSRCGLVLTLHLPAEEAFPFLRGLTPLTADRLSTYL
                DRAGALRSLTPLVELLTLSAKKQPQGDARGRVAWLRPKIVGCLRRIYRVNISARWFIS
                TFGSHEAQFVLVTAAYYFWGIPCTIETLAHLTELFTSESGQSLAAVTSLAELGEVFGS
                SAWAEQTEAFAHFAHEKLRRDSREIRAVARTIDAYRGRLPLASADLVRYVYLAHAQCF
                NEGTFKRYSQLTSMGEIGCLPSGGVVLPSLLDRGFAEHMRTYFTRETYLAEHVRVQQL
                KIRMEPPAPYTWDPDPDDGLMRAWAGLSVDVARELVELARWHADEGPTYPPTLQGFLC
                LAGQATCRGQWNPKEQFLPPTVLRRVQRLPVFLCHFADRHYFVMTAADPFSSHLAEVV
                STPTNCRLPDTCLTRALSYTPVYYSQNSLSEQLFVSRHEYFNPRLPVCNLVLDLDLKI
                KGAPWSLEEIYDLCRTVRREVLRLMRRLGPVSRAHPVYFFKSACPPADPDNMEDVLPF
                CICTGKLGFRVITPLPRGHAIVGTSAVQGFVSVLQKLMGLTACLRRMRHKIKEIGAPL
                FDSGVYHAGRCIRLPHTYKVDRGGGLSRQLRLFVCHPEEEDKHSYVKNALNIQNLLHH
                SLHVGWPAPKTFCYHIADDGRDYLIQRTRETLPPTVENVCAMIEGHLGLDLVAWVSSC
```

```
                     IWPSLMSTLATAVPEDKFPQFLHVTFEQTGPNLVQVCHARGRNFACLRHTHRASSKNV
                     RVFLVLYYTSQAITVTFMSQCFAGRCGANQPTAHFSISVPASRIINRAEASQDSTTSQ
                     LARRRDRQDGSFSETLPN"
promoter             complement(84356)
                     /note="TATA: CATAAAT before BSLF2 and BMLF1. Two RNAs
                     start here; one is spliced and the other is unspliced,
                     both traverse BMLF1."
promoter             86882
                     /note="TATA: TATTTAA BS-R1 late promoter before BSRF1"
CDS                  86924..87580
                     /note="BSRF1 reading frame;  NCBI gi: 59104"
                     /codon_start=1
                     /translation="MAFYLPDWSCCGLWLFGRPRNRYSQLPEEPETFECPDRWRAEID
                     LGLPPGVQVGDLLRNEQTMGSLRQVYLLAVQANSITDHLKRFDAVRVPESCRGVVEAQ
                     VAKLEAVRSVIWNTMISLAVSGIEMDENGLKALLDKQAGDSLALMEMEKVATALKMDE
                     TGAWAQEISAVVSSVTAPSASAPFINSAFEPEVPTPVLAPPPVVRQPEHSGPTELALT
                     "
misc_feature         complement(87134)
                     /note="polyA signal: AATAAA"
misc_feature         87599
                     /note="polyA signal: AATAAA"
misc_feature         complement(87613)
                     /note="polyA signal: AATAAA, end 1.0kb early RNA from
                     BLL3"
CDS                  complement(87638..88474)
                     /note="BLLF3 early reading frame (BLLF2 in Baer et al,
                     1984). Homologous to RF 8 VZV and dUTPase HSV.  NCBI gi:
                     59105"
                     /codon_start=1
                     /translation="MEACPHIRYAFQNDKLLLQQASVGRLTLVNKTTILLRPMKTTTV
                     DLGLYARPPEGHGLMLWGSTSRPVTSHVGIIDPGYTGELRLILQNQRRYNSTLRPSEL
                     KIHLAAFRYATPQMEEDKGPINHPQYPGDVGLDVSLPKDLALFPHQTVSVTLTVPPPS
                     IPHHRPTIFGRSGLAMQGILVKPCRWRRGGVDVSLTNFSDQTVFLNKYRRFCQLVYLH
                     KHHLTSFYSPHSDAGVLGPRSLFRWASCTFEEVPSLAMGDSGLSEALEGRQGRGFGSS
                     GQ"
misc_feature         87650
                     /note="BAM: Bam H1 S/L"
promoter             88507
                     /note="TATA: TATATAT BL-R1 late promoter before BLRF1,
                     gives 1.0kb late RNA"
promoter             88511
                     /note="TATA: TATAAGA"
promoter             complement(88514)
                     /note="TATA: TATATAT BL-L3 early promoter before BLLF3,
                     gives 1.0kb early RNA"
CDS                  88547..88855
                     /note="BLRF1 late reading frame;  NCBI gi: 59106"
                     /codon_start=1
                     /translation="MGKVLRKPFAKAVPLLFLAATWLLTGVLPAGASSPTNAAAASLT
                     EAQDQFYSYTCNADTFSPSLTSFASIWALLTLVLVIIASAIYLMYVCFNKFVNTLLTD
                     "
promoter             88863
                     /note="TATA: TATTTAA BL-R2 late promoter before BLRF2,
                     gives 0.6kb late RNA"
CDS                  88925..89413
                     /note="BLRF2 late reading frame, 2 NXS/T;  NCBI gi: 59107"
                     /codon_start=1
                     /translation="MSAPRKVRLPSVKAVDMSMEDMAARLARLESENKALKQQVLRGG
                     ACASSTSVPSAPVPPPEPLTARQREVMITQATGRLASQAMKKIEDKVRKSVDGVTTRN
                     EMENILQNLTLRIQVSMLGAKGQPSPGEGTRPRESNDPNATRRARSRSRGREAKKVQI
```



```
                     SD"
misc_feature         89412
                     /note="polyA signal: AATAAA, end of 1.0kb and 0.6kb late
                     RNAs"
misc_feature         complement(89425)
                     /note="polyA signal: AATAAA, end of 0.7kb early, 2.2kb
                     late and 2.8kb late RNA"
CDS                  complement(89430..92153)
                     /note="BLLF1a, late reading frame, gp350 membrane antigen,
                     36 NXT/S (Hummel et al, 1984; Biggin et al, 1984; Beisel
                     et al, 1985);  NCBI gi: 59110"
                     /codon_start=1
                     /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                     DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                     ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
                     STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
                     PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
                     ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
                     NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
                     KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
                     PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
                     SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
CDS                  complement(89430..92153)
                     /note="BLLF1b, late reading frame gp220 membrane antigen,
                     spliced form of BLLF1a (Hummel et al, 1984; Biggin et al,
                     1984; Beisel et al, 1985);  NCBI gi: 59109"
                     /codon_start=1
                     /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                     DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                     ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
                     STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
                     PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
                     ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
                     NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
                     KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
                     PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
                     SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
promoter             complement(89434)
                     /note="TATA: TATAAAG"
CDS                  complement(89567..90013)
                     /note="BLLF2 early reading frame (BLLF3 in Baer et al,
                     1984);  NCBI gi: 59108"
                     /codon_start=1
                     /translation="MCPPVRQHPAQAPPAKRQALETVPHPQNRGRLMSPKARPPKMQR
                     RPRPPVAKRRRFPRSPQQVERPILPPVESTPQDMEPGQVQSPPQITAVIQLRQDRDTM
                     RPPIYLPALLANCGPAGLLRAHRLPQPKPPCQSRQRPSPDSQTSPC"
promoter             complement(90051)
                     /note="TATA: TATAACA BL-L2 early promoter before BLLF2,
```

```
                gives 0.7kb early RNA"
intron          complement(90062..90652)
                /note="intervening sequence in gp220 gene"
repeat_region   90177..90639
                /note="21 copies of 21bp approximate repeat"
promoter        complement(92192)
                /note="TATA: TATTAAA BL-L1 late promoter before BLLF1a,b.
                Gives 2.8 and 2.2kb late RNAs"
mRNA            92238..92581
                /note="Exon in Bodescot et al (1986) RNA (spliced from
                20763 to 92670)"
CDS             92243..92581
                /note="BLRF3 reading frame;  NCBI gi: 59111"
                /codon_start=1
                /translation="MDKDRPGPPALDDNMEEEVPSTSVVQEQVSAGDWENVLIELSDS
                SSEKEAEDAHLEPAQKGTKRKRVDHDAGGSAPARPMLPPQPDLPGREAILRRFPLDLR
                TLLQAIGAAAT"
CDS             92670..95165
                /note="BERF1 frame, homology with BERF2b and BERF4. A
                fusion of BLRF3 with BERF1 encodes EBNA-3A, latent cycle
                gene. (Hennessy et al, 1986, Joab et al, 1987);  NCBI gi:
                583889"
                /codon_start=1
                /translation="XIDTRAIDQFFGSQISNTEMYIMYAMAIRQAIRDRRRNPASRRD
                QAKWRLQTLAAGWPMGYQAYSSWMYSYTDHQTTPTFVHLQATLGCTGGRRCHVTFSAG
                TFKLPRCTPGDRQWLYVQSSVGNIVQSCNPRYSIFFDYMAIHRSLTKIWEEVLTPDQR
                VSFMEFLGFLQRTDLSYIKSFVSDALGTTSIQTPWIDDNPSTETAQAWNAGFLRGRAY
                GIDLLRTEGEHVEGATGETREESEDTESDGDDEDLPCIVSRGGPKVKRPPIFIRRLHR
                LLLMRAGKRTEQGKEVLEKARGSTYGTPRPPVPKPRPEVPQSDETATSHGSAQVPEPP
                TIHLAAQGMAYPLHEQHGMAPCPVAQAPPTPLPPVSPGDQLPGVFSDGRVACAPVPAP
                AGPIVRPWEPSLTQAAGQAFAPVRPQHMPVEPVPVPTVALERPVYPKPVRPAPPKIAM
                QGPGETSGIRRRARERWRPAPWTPNPPRSPSQMSVRDRLARLRAEAQVKQASVEVQPPQ
                LTQVSPQQPMEGPLVPEQQMFPGAPFSQVADVVRAPGVPAMQPQYFDLPLIQPISQGA
                PVAPLRASMGPVPPVPATQPQYFDIPLTEPINQGASAAHFLPQQPMEGPLVPEQWMFP
                GAALSQSVRPGVAQSQYFDLPLTQPINHGAPAAHFLHQPPMEGPWVPEQWMFQGAPPS
                QGTDVVQHQLDALGYTLHGLNHPGVPVSPAVNQYHLSQAAFGLPIDEDESGEGSDTSE
                PCEALDLSIHGRPCPQAPEWPVQEEGGQDATEVLDLSIHGRPRPRTPEWPVQGEGGQN
                VTGPETRRVVVSAVVHMCQDDEFPDLQDPPDEA"
mRNA            92670..95248
                /note="Exon in (Bodescot et al, 1986) RNA from 92581, to
                3' end"
misc_feature    92703
                /note="BAM: Bam H1 L/E"
promoter        complement(93161)
                /note="TATA: CATAAAT"
promoter        93479
                /note="TATA: TATAAGA"
promoter        complement(93482)
                /note="TATA: TATAAAT"
repeat_region   94208..94277
                /note="repeat type A"
repeat_region   94281..94306
                /note="repeat type B"
repeat_region   94307..94381
                /note="repeat type C"
repeat_region   94386..94411
                /note="repeat type B"
repeat_region   94412..94489
                /note="repeat type C"
repeat_region   94490..94560
```



```
                /note="repeat type A"
repeat_region   94571..94648
                /note="repeat type C"
repeat_region   94649..94719
                /note="repeat type A"
repeat_region   94896..94982
                /note="repeat type D"
repeat_region   94983..95069
                /note="repeat type D"
misc_feature    95221
                /note="polyA signal: AATAAA"
misc_feature    complement(95272)
                /note="polyA signal: AATAAA"
CDS             95353..95724
                /note="BERF2a reading frame;  NCBI gi: 59113"
                /codon_start=1
                /translation="MKKAWLSRAQQADAGGASGSEDPPDYGDQGNVTQVGSEPISPEI
                GPFELSAASEDDPQSGPVEENLDAAAREEEEPHEQEHNGGDDPLDVHTRQPRFVDVNP
                TQAPVIQLVHAVYDSMLVRGT"
CDS             95725..98247
                /note="BERF2b frame, homology with BERF1 and BERF4. BERF2a
                and BERF2b are spliced together to make EBNA3B (EBNA4A)
                latent protein.  NCBI gi: 583890"
                /codon_start=1
                /translation="XISRCFAWIFWPVLIDCHWFQQSDLRPLGSLFLEQNLNIEEFIW
                MCMTVRHRCQAIRKKPLPIVKQRRWKLLSSCRSWRMGYRTHNLKVNSFESGGDNVHPV
                LVTATLGCDEGTRHATTYSAGIVQIPRISDQNQKIETAFLMARRARSLSAERYTLFFD
                LVSSGNTLYAIWIGLGTKNRVSFIEFVGWLCKKDHTHIREWFRQCTGRPKAAKPWLRA
                HPVAIPYDDPLTNEEIDLAYARGQAMNIEAPRLPDDPIIVEDDDESEEIEAESDEEED
                KSGMESLKNIPQTLPYNPTVYGRPAVFDRKSDAKSTKKCRAIVTDFSVIKAIEEEHRK
                KKAARTEQPRATPESQAPTVVLQRPPTQQEPGPVGPLSVQARLEPWQPLPGPQVTAVL
                LHEESMQGVQVHGSMLDLLEKDDEVMEQRVMATLLPPVPQQPRAGRRGPCVFTGDLGI
                ESDEPASTEPVHDQLLPAPGPDPLEIQPLTSPTTSQLSSSAPSCAQTPWPVVQPSQTP
                DDPTKQSRPPETAAPRQWPMPLRPIPMRPLRMQPIPFNHPVGPTPHQTPQVEITPYKP
                TWAQIGHIPYQPTPTGPATMLLRQWAPATMQTPPRAPTPMSPPEVPPVPRQRPRGAPT
                PTPPPQVPPVPRQRPRGAPTPTPPPQVLPTPMQLALRAPAGQQGPTKQILRQLLTGGV
                KKGRPSLKLQAALERQAAAGWQPSPGSGTSDKIVQAPIFYPPVLQPIQVMGQGGSPTA
                MAASAVTQAPTEYTRERRGVGPMPPTDIPPSKRAKIEAYTEPEMPHGGASHSPVVILE
                NVGQGQQQTLECGGTAKQERDMLGLGDIAVSSPSSSETSNDE"
misc_feature    complement(95819)
                /note="polyA signal: AATAAA"
promoter        complement(95853)
                /note="TATA: TATAAAT"
misc_feature    complement(96276)
                /note="polyA signal: AATAAA"
repeat_region   97522..97698
                /note="3x60bp repeat"
CDS             98323..98769
                /note="BERF3 reading frame;  NCBI gi: 59115"
                /codon_start=1
                /translation="MTILDEVEILHIFRPTMESFEGQGDSRQSPDNERGDNVQTTGEH
                DQDPGPGPPSSGASERLVPEESYSRDQQPWGQSRGDENRGWMQRIRRRRRRRAALSGH
                LLDTEDNVPPWLPPHDITPYTARNIRDAACRAVKVSMPLTGFMGAI"
mRNA            98364..98730
                /note="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"
misc_feature    98731
                /note="DONOR: AAGGTGAGT donor"
CDS             98805..101423
                /note="BERF4 frame, homology with BERF1 and BERF2b. BERF3
                and BERF4 are spliced together to make the EBNA3C (EBNA
```

```
                     4B) latent protein.  NCBI gi: 583891"
                     /codon_start=1
                     /translation="XSHLQALSNLILDSGLDTQHILCFVMAARQRLQDIRRGPLVAEG
                     GVGWRHWLLTSPSQSWPMGYRTATLRTLTPVPNRVGADSIMLTATFGCQNAARTLNTF
                     SATVWTPPHAGPREQERYAREAEVRFLRGKWQRRYRRIYDLIELCGSLHHIWQNLLQT
                     EENLLDFVRFMGVMSSCNNPAVNYWFHKTIGNFKPYYPWNAPPNENPYHARRGIKEHV
                     IQNAFRKAQIQGLSMLATGGEPRGDATSETSSDEDTGRQGSDVELESSDDELPYIDPN
                     MEPVQQRPVMFVSRVPAKKPRKLPWPTPKTHPVKRTNVKTSDRSDKAEAQSTPERPGP
                     SEQSSVTVEPAHPTPVEMPMVILHQPPPVPKPVPVKPTPPPSRRRRGACVVYDDDVIE
                     VIDVETTEDSSSVSQPNKPHRKHQDGFQRSGRRQKRAAPPTVSPSDTGPPAVGPPAAG
                     PPAAGPPAAGPPAAGPPAAGPPAAGPRILAPLSAGPPAAGPHIVTPPSARPRIMAPPV
                     VRMFMRERQLPQSTGRKPQCFWEMRAGREITQMQQEPSSHLQSATQPTTPRPSWAPSV
                     CALSVMDAGKAQPIESSHLSSMSPTQPISHEEQPRYEDPDAPLDLSLHPDVAAQPAPQ
                     APYQGYQEPPAPQAPYQGYQEPPPPQAPYQGYQEPPAHGLQSSSYPGYAGPWTPRSQH
                     PCYRHPWAPWSQDPVHGHTQGPWDPRAPHLPPQWDGSAGHGQDQVSQFPHLQSETGPP
                     RLQLSLVPLVSSSAPSWSSPQPRAPIRPIPTRFPPPPPMPLQDSMAVGCDSSGTACPSM
                     PFASDYSQGAFTPLDINATTPKRPRVEESSHGPARCSQATAEAQEILSDNSEISVFPK
                     DAKQTDYDASTESELD"
mRNA                 98805..99050
                     /note="Exon in T4 cDNA (Bodescot et al 1986). 99050 is not
                     the end of the RNA."
misc_feature         99126..102118
                     /note="DEL: Deletion in Raji"
promoter             99443
                     /note="TATA: CATAAAA"
misc_feature         100104
                     /note="DONOR: ACCGTGAGT possible donor before repeat."
repeat_region        100122..100304
                     /note="10 x 15bp repeat"
misc_feature         100528
                     /note="DONOR: CTGGTAAGG possible donor"
misc_feature         100613
                     /note="BAM: Bam H1 E/e1"
repeat_region        100665..100781
                     /note="3x39bp repeat"
promoter             complement(100860)
                     /note="TATA: TATAACA"
misc_feature         100919
                     /note="BAM: Bam H1 e1/e2"
misc_feature         101426
                     /note="BAM: Bam H1 e2/e3"
CDS                  complement(101445..102116)
                     /note="BZLF2 reading frame 3x NXT/S. 2.5kb late RNA
                     traverses BZLF2, ends unknown.  NCBI gi: 59117"
                     /codon_start=1
                     /translation="MVSFKQVRVPLFTAIALVIVLLLAYFLPPRVRGGGRVAAAAITW
                     VPKPNVEVWPVDPPPPVNFNKTAEQEYGDKEVKLPHWTPTLHTFQVPQNYTKANCTYC
                     NTREYTFSYKGCCFYFTKKKHTWNGCFQACAELYPCTYFYGPTPDILPVVTRNLNAIE
                     SLWVGVYRVGEGNWTSLDGGTFKVYQIFGSHCTYVSKFSTVPVSHHECSFLKPCLCVS
                     QRSNS"
promoter             101690
                     /note="TATA: CATAAAA"
misc_feature         101765
                     /note="polyA signal: AATAAA"
promoter             complement(101786)
                     /note="TATA: TATAAAG"
misc_feature         101947
                     /note="BAM: Bam H1 e3/Z"
misc_feature         complement(102098)
                     /note="DONOR: CAGGTGAGG possible donor"
```



```
mRNA                 complement(102126..102341)
                     /note="3' terminal exon of 0.9kb and 2.8kb early RNAs"
promoter             102153
                     /note="TATA: TATTAAT"
misc_feature         complement(102156)
                     /note="polyA signal: AATAAA 3' end of 0.9kb and 2.8kb RNAs
                     encoding BZLF1 and BRLF1"
promoter             complement(102160)
                     /note="TATA: TATTAAT"
CDS                  complement(102213..103155)
                     /note="BZLF1 reading frame, modified from Baer et al,
                     1984. Has two splices within frame. 2xNXT/S. Immediate
                     early gene which disrupts latency (Countryman and Miller,
                     1985), called EB1 by Chevallier-Greco et al, 1986 and
                     ZEBRA by Miller."
                     /codon_start=1
promoter             complement(102380)
                     /note="TATA: CATAAAT"
promoter             102415
                     /note="TATA:TATATAC"
promoter             complement(102420)
                     /note="TATA: TATATAC"
mRNA                 complement(102426..102530)
                     /note="Exon of 0.9kb and 2.8kb early RNAs"
misc_feature         complement(102504)
                     /note="polyA signal: AATAAA, apparently not functional"
repeat_region        102581..102652
                     /note="semi-repetitive sequence, homologous to human c-fos
                     3' sequence"
mRNA                 complement(102655..103194)
                     /note="First exon of 0.9kb early RNA encoding BZLF1"
misc_feature         complement(102918)
                     /note="splice acceptor used in RZ fusion gene (Sargeant)"
promoter             complement(103231)
                     /note="TATA: TTTAAA of BZL1 immediate early promoter gives
                     0.9kb RNA"
misc_feature         complement(103256..103311)
                     /note="Upstream of BZL1, homology to 106243 to 106188"
CDS                  complement(103366..105183)
                     /note="BRLF1 reading frame, (immediate?) early gene, acts
                     as transcription activator.  NCBI gi: 59119"
                     /codon_start=1
                     /translation="MRPKKDGLEDFLRLTPEIKKQLGSLVSDYCNVLNKEFTAGSVEI
                     TLRSYKICKAFINEAKAHGREWGGLMATLNICNFWAILRNNRVRRRAENAGNDACSIA
                     CPIVMRYVLDHLIVVTDRFFIQAPSNRVMIPATIGTAMYKLLKHSRVRAYTYSKVLGV
                     DRAAIMASGKQVVEHLNRMEKEGLLSSKFKAFCKWVFTYPVLEEMFQTMVSSKTGHLT
                     DDVKDVRALIKTLPRASYSSHAGQRSYVSGVLPACLLSTKSKAVETPILVSGADRMDE
                     ELMGNDGGASHTEARYSESGQFHAFTDELESLPSPTMPLKPGAQSADCGDSSSSSSDS
                     GNSDTEQSEREEARAEAPRLRAPKSRRTSRPNRGQTPCPSNAAEPEQPWIAAVHQESD
                     ERPIFPHPSKPTFLPPVKRKKGLRDSREGMFLPKPEAGSAISDVFEGREVCQPKRIRP
                     FHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDP
                     DEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNL
                     DSPLTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF"
misc_feature         complement(103453..103462)
                     /note="TAATGAAATC sequence"
misc_feature         103741
                     /note="BAM: Bam H1 Z/g"
misc_feature         103816
                     /note="BAM: Bam H1 g/R"
mRNA                 complement(104926..105185)
```

```
                /note="exon in RZ fusion gene (Sargeant)"
mRNA            complement(104927..104989)
                /note="BRLF2 poss. small 5' exon"
promoter        105016
                /note="TATA: TATAAAT before BRRF1, possible promoter for
                1.1 kb early RNA encoding BRRF1"
promoter        complement(105019)
                /note="TATA: TATAAAT before BRLF2"
CDS             105182..106114
                /note="BRRF1 early reading frame;  NCBI gi: 59118"
                /codon_start=1
                /translation="MASSNRGNARPLKSFLHELYLKHYPEVGDVVHLLNTIGVDCDLP
                PSHPLLTAQRGLFLARVLQAVQQHKLLEDTIVPKILKKLAYFLELLSYYSPKDEQRDI
                AEVLDHLKTNRDLGLDDRLWALIRKLRQDRHHASVNVLMPGSDYTAVSLQYYDGISIG
                MRKVIADVCRSGYASMPSMTATHNLSHQLLMASGPSEEPCAWRGFFNQVLLWTVALCK
                FRRCIYYNYIQGSIATISQLLHLEIKALCSWIISQDGMRLFQHSRPLLTLWESVAANQ
                EVTDAITLPDCAEYIDLLKHTKHVLENCSAMQYK"
misc_feature    complement(105185)
                /note="ACCEPT: splice acceptor in 2.8kb early RNA encoding
                BRLF1 and RZ fusion gene (Sargeant)"
promoter        105213
                /note="TATA: CATTAAA"
misc_feature    106110
                /note="polyA signal: AATAAA, 3' end of early 1.1kb RNA
                encoding BRRF1"
misc_feature    complement(106125)
                /note="DONOR: CAGGTAAGA possible donor"
misc_feature    complement(106188..106243)
                /note="Homology to upstream region of BZL1"
promoter        complement(106213)
                /note="TATA: CATAAAA"
promoter        106243
                /note="TATA: TATAAAA before BRRF2, possible promoter for
                1.8 kb RNA encoding BRRF2"
CDS             106302..107915
                /note="BRRF2 reading frame;  NCBI gi: 59120"
                /codon_start=1
                /translation="MSGQQRGSVILVPEHLAGALTKLMSDFITGQDVTLSGGNIAVKI
                RDAINQTPGGGDVAILSSLFALWNALPTSGRQSSRDDLIPAAVQALTTAHNLCLGVIP
                GETSHKDTPESLLRAIVTGLQKLWVDSCGCPECLQCLKGLKAIKPGLYEIPRIIPHTK
                QCSPVNLLNMLVHKLVALRGHVQLAYDARVLTPDFHEIPDLDDSDAVFARTLLAALFH
                LNMFFILKDYITQDSMSLKQALSGHWMSATGNPLPAAPETLRDYLEAFRNSDNHFYLP
                TTGPLNTFQFPEELLGRVVVIDSSLCAASHVQDVITHGVGAGVPRPRFSALPPAPSRE
                PQQTCSQLTSRGNESSRRNLGQPGGTSPAVPPVCPIVSLTASGAKQNRGGMGSLHLAK
                PEETSPAVSPVCPIASPAASRSKQHCGVTGSSQAAPSFSSVAPVASLSGDLEEEEEGS
                RESPSLPSSKKGDEEFEAWLEAQDANLEDVQREFSGLRVIGDEDEDGSEDGEFSDLDL
                SDSDHEGDEGGGAVGGGRSLHSLYSLSVV"
promoter        complement(106385)
                /note="TATA: GATAAAA"
misc_feature    complement(106973)
                /note="polyA signal: AATAAA"
promoter        complement(107124)
                /note="TATA: GATAAAA"
misc_feature    107457
                /note="BAM: Bam H1 R/f"
misc_feature    107565
                /note="BAM: Bam H1 f/K"
misc_feature    107914
                /note="polyA signal: AATAAA, 3' end of 1.8kb RNA encoding
                BBRF2"
```



```
misc_feature    107942
                /note="ACCEPT: splice acceptor for EBNA-1 RNA (from
                98730)"
CDS             107950..109875
                /note="BKRF1 encodes EBNA-1 protein, latent cycle gene.
                NCBI gi: 59121"
                /codon_start=1
                /translation="MSDEGPGTGPGNGLGEKGDTSGPEGSGGSGPQRRGGDNHGRGRG
                RGRGRGGGRPGAPGGSGSGPRHRDGVRRPQKRPSCIGCKGTHGGTGAGAGAGGAGAGG
                AGAGGGAGAGGGAGGAGGAGGAGAGGGAGAGGGAGGAGGAGAGGGAGAGGGAGGAGAG
                GGAGGAGGAGAGGGAGAGGGAGGAGAGGGAGGAGGAGAGGGAGAGGAGGAGGAGAGGA
                GAGGGAGGAGGAGAGGAGAGGAGAGGAGAGGAGGAGAGGAGAGGAGAGGAGAGGAGAG
                GAGAGGGAGGAGAGGAGGAGAGGAGGAGAGGAGGAGAGGGAGAGGAGAGGGGRGRGGS
                GGRGRGGSGGRGRGGSGGRRGRGRERARGGSRERARGRGRGRGEKRPRSPSSQSSSSG
                SPPRRPPPGRRPFFHPVGEADYFEYHQEGGPDGEPDVPPGAIEQGPADDPGEGPSTGP
                RGQGDGGRRKKGGWFGKHRGQGGSNPKFENIAEGLRALLARSHVERTTDEGTWVAGVF
                VYGGSKTSLYNLRRGTALAIPQCRLTPLSRLPFGMAPGPGPQPGPLRESIVCYFMVFL
                QTHIFAEVLKDAIKDLVMTKPAPTCNIRVTVCSFDDGVDLPPWFPPMVEGAAAEGDDG
                DDGDEGGDGDEGEEGQE"
repeat_region   108217..108924
                /note="EBNA triplet repeat GGA,GCA,GGG."
misc_feature    109856
                /note="DONOR: AGGGTGAGG possible donor at end BKRF1"
promoter        109905
                /note="TATA: TATTAAA before BKRF2, possible start for
                2.3kb late RNA"
misc_feature    109906
                /note="polyA signal: ATTAAA"
misc_feature    109937
                /note="polyA signal: AATAAA 3' end of EBNA-1 RNA"
CDS             109958..110371
                /note="BKRF2 reading frame;  NCBI gi: 59122"
                /codon_start=1
                /translation="MRAVGVFLAICLVTIFVLPTWGNWAYPCCHVTQLRAQHLLALEN
                ISDIYLVSNQTCDGFSLASLNSPKNGSNQLVISRCANGLNVVSFFISILKRSSSALTG
                HLRELLTTLETLYGSFSVEDLFGANLNRYAWHRGG"
misc_feature    110271
                /note="DONOR: TCCGTGAGT possible donor at end of BKRF2"
CDS             110275..111120
                /note="BKRF3 reading frame, homologous to RF 59 VZV;  NCBI
                gi: 583892"
                /codon_start=1
                /translation="XVNHPGDSLRFILSGRPVWCQLKQIRMASRGLDLWLDEHVWKRK
                QEIGVKGENLLLPDLWLDFLQLSPIFQRKLAAVIACVRRLRTQATVYPEEDMCMAWAR
                FCDPSDIKVVILGQDPYHGGQANGLAFSVAYGFPVPPSLRNIYAELHRSLPEFSPPDH
                GCLDAWASQGVLLLNTILTVQKGKPGSHADIGWAWFTDHVISLLSERLKACVFMLWGA
                KAGDKASLINSKKHLVLTSQHPSPLAQNSTRKSAQQKFLGNNHFVLANNFLREKGLGE
                IDWRL"
misc_feature    111098
                /note="DONOR: TCGGTGAGA possible donor at end BKRF3"
CDS             111107..111787
                /note="BKRF4 reading frame, contains complex repetitive
                sequence;  NCBI gi: 583893"
                /codon_start=1
                /translation="XGGCRGVITMAMFLKSRGVRSCRDRRLLSDEEEETSQSSSYTLG
                SQASQSIQEEDVSDTDESDYSDEDEEIDLEEEYPSDEDPSEGSDSDPSWHPSDSDESD
                YSESDEDEATPGSQASRSSRVSPSTQQSSGLTPTPSFSRPRTRAPPRPPAPAPVRGRA
                SAPPRPPAPVQQSTKDKGPHRPTRPVLRGPAPRRPPPPSSPNTYNKHMMETTPPIKGN
                NNYNWPWL"
misc_feature    111272
```

```
                     /note="DONOR: GACGTGAGT poss.donor before rpt.seq. in
                     BKRF4"
     misc_feature    111719
                     /note="polyA signal: AATAAA"
     misc_feature    111787
                     /note="polyA signal: AATAAA : currently unknown which is
                     3' end of the 2.3kb late and 1.1kb early RNAs"
     CDS             complement(111830..114259)
                     /note="BBLF4 early reading frame, very good homology to
                     RF55 VZV;  NCBI gi: 59126"
                     /codon_start=1
                     /translation="MAEEPRAPEALSSTFMLNMTSDASVRRIVRRIGTLARRRVQQLP
                     DMETFSPEFDPELSEPPFLPFSAYVITGTAGAGKSTSVSCLHHTMDCLVTGATTVAAQ
                     NLSQTLRAYCPTVYSAFGFKSRHINMTQRVSSHGRSTDAALEELQRRDLAKYWPVLSD
                     IAAEFRRTKPRGLYSGVSGPAFEVLRDMHQGQLWTTNVIVVDEAGTLSVHILTAVVFC
                     YWFFNAWLRTPLYRRGRIPCIVCVGSPTQTDAFQSSFSHETQVNKIRECDNILTFLVG
                     NPRAATYVDVARNWALFINNKRCTDVQFGHLMKTLEYGLELSPDILAYVDRFVVPRAA
                     IMDPAQYVGWTRLFLSHAEVKTFLTTLHATLKTAGQGRAARGTGGDGGGVTMFTCPVE
                     CEVFLDPLAQYKTLVGLPGLTAHTWLQKNYARLGNYSQFADQDMVPVGTEQDEERVKV
                     TYNVTYVKHSSVSVNCKTKKSICGYTGTFGDFMDTLEADSFVEAHGHEQPEYVYSFLA
                     RLIYGGIYAFSHGGHSLCENGEYVAELGAVPLPGRTWDPEVTAGMELGELPLEVAWDG
                     ERSPAAVFYARVLAPPAANSAPLCSLLNIYNDLRAYFRQCLDVAVRYGGREFRDLPFC
                     TFTNNMLIRDNIEFTSDEPLLHGLLDYASTTENYTLLGYTHLNVFFGIRGKQQPQDAG
                     SSRMPRLMVKDEAGFVCCLEHNTNKLYETIEDKSLNLCSIRDYGISSKLAMTIAKAQG
                     LSLNKVAICFGSHRNIKPGHVYVALSRARHSNCVVMDRNPLSEMITGEGNPASGYIVD
                     ALKNSRALLVY"
     misc_feature    complement(111830)
                     /note="polyA signal: AATAAA"
     promoter        112471
                     /note="TATA: TATATAT"
     promoter        complement(112476)
                     /note="TATA: TATATAA"
     misc_feature    112620
                     /note="BAM: Bam H1 K/B"
     promoter        113876
                     /note="TATA: TATTTAT before BBRF1"
     promoter        complement(113885)
                     /note="TATA: CATAAAT"
     CDS             114204..116045
                     /note="BBRF1 late reading frame, homologous to RF 54 VZV;
                     NCBI gi: 59125"
                     /codon_start=1
                     /translation="MFNMNVDESASGALGSSAIPVHPTPASVRLFEILQGKYAYVQGQ
                     TIYANLRNPGVFSRQVFTHLFKRAISHCTYDDVLHDWNKFEACIQKRWPSDDSCASRF
                     RESTFESWSTTMKLTVRDLLTTNIYRVLHSRSVLSYERYVDWICATGMVPAVKKPITQ
                     ELHSKIKSLRDRCVCRELGHERTIRSIGTELYEATKEIIESLNSTFIPQFTEVTIEYL
                     PRSDEYVAYYCGRRIRLHVLFPPAIFAGTVTFDSPVQRLYQNIFMCYRTLEHAKICQL
                     LNTAPLKAIVGHGGRDMYKDILAHLEQNSQRKDPKKELLNLLVKLSENKTISGVTDVV
                     EEFITDASNNLVDRNRLFGQPGETAAQGLKKKVSNTVVKCLTDQINEQFDQINGLEKE
                     RELYLKKIRSMESQLQASLGPGGNNPAASAPAAVAAEAASVDILTGSTASAIEKLFNS
                     PSASLGARVSGHNESILNSFVSQYIPPSREMTKDLTELWESELFNTFKLTPVVDNQGQ
                     RLYVRYSSDTISILLGPFTYLVAELSPVELVTDVYATLGIVEIIDELYRSSRLAIYIE
                     DLGRKYCPASATGGDHGIRQAPSARGDTEPDHAKSKPARDPPPGAGS"
     CDS             115843..116784
                     /note="BBRF2 late reading frame, homologous to RF 53 VZV;
                     NCBI gi: 583894"
                     /codon_start=1
                     /translation="XSWRSSTSSTGAVAWPSTSRTSVENTAPRARPGEIMASGKHHQP
                     GGTRSLTMQKVSLRVTPRLVLEVNRHNAICVATNVPEFYNARGDLNIRDLRAHVKARM
                     ISSQFCGYVLVSLLDSEDQVDHLNIFPHVFSERMILYKPNNVNLMEMCALLSMIENAK
```



```
                     SPSIGLCREVLGRLTLLHSKCNNLDSLFLYNGARTLLSTLVKYHDLEEGAATPGPWNE
                     GLSLFKLHKELKRAPSEARDLMQSLFLTSGKMGCLARSPKDYCADLNKEEDANSGFTF
                     NLFYQDSLLTKHFQCQTVLQTLRRKCLGSDTVSKIIP"
     promoter        complement(116683)
                     /note="TATA: GATAAAA"
     misc_feature    complement(116696)
                     /note="polyA signal: AATAAA"
     CDS             complement(116781..117386)
                     /note="BBLF3 early reading frame, spliced to BBLF2. BBLF3
                     contains a consensus nucleotide binding site;  NCBI gi:
                     583895"
                     /codon_start=1
                     /translation="XFLQGVKDSEDASRLDRDVMGGEATVARRHIRVKARRGPGCLLM
                     AIFQGDLYVGGCREHSGPFLVWHEAFSWTLDQLAARPEADKAPPSHDHLLTLVRDLTR
                     RLAPGRRRNRFWALPRAWLQRLRRAGLRLSGSHVCLLDKDGARPAPCQTATEHGLSPT
                     AYFREIMAFLLDVISALHPGYTIPMEITRETDLLMTVLSLF"
     misc_feature    116785
                     /note="polyA signal: AATAAA"
     intron          complement(117386..117515)
                     /note="intron spliced out in RNA linking BBLF2 and BBLF3"
     CDS             complement(117515..119080)
                     /note="BBLF2 early reading frame, spliced to BBLF3;  NCBI
                     gi: 583896"
                     /codon_start=1
                     /translation="XDVLKGPVLLRSQTMMETPAESVRARVSSVTFYNVTQTAGRWWA
                     IWVVGIVPIKREDVETLIVVQACQPPLGGSLEPPVVNAPSTTELNFLRWERELRRSGG
                     LIAMLADAAEKDLFDLSFRTRDRRLLSAARVEDEQGLIFQPLFPAQVVCQSCSGDDGR
                     DQQPPPVDGFGSEMEGEQTCPHAQRHSESPGQLDVYIRTPRGDVFTYSTETPDDPSPV
                     PFRDILRPVTYEVDLVSSDGATGRGGDARRHRVSLKILEPAGGFESWLVNSWSMAGGG
                     LYAFLRSIYASCYANHRGTKPIFYLLDPELCPGGSDFQPYVPGFPFLPIHYVGRARPA
                     FWHRAPHSEGLLLLDLNLGVSGTPLADALLGLDARSGQRRGSLLLQQIWPPTRKEINP
                     RHVCTREGGEGGGEDETTVVGRAEATAILEADATWWLYELARCHLSARGAPVGTPDGG
                     GQARDAQTWLRALHRYGTSDTRRALGGLYTAVTRVLLHAAADLGLTWAYADEFILGFV
                     APTSAHPSEEPLAQ"
     promoter        118981
                     /note="TATA: TATAAAA BBR1 late promoter before BBRF3"
     promoter        119067
                     /note="TATA: TTTAAAA BBR2 late promoter ?"
     promoter        119098
                     /note="TATA: TATTTAA BBR3 late promoter before BBRF3"
     misc_feature    119108
                     /note="DONOR: AAGGTGAAT possible donor"
     CDS             119137..120354
                     /note="BBRF3 late reading frame;  NCBI gi: 59130"
                     /codon_start=1
                     /translation="MKSSKNDTFVYRTWVKTLVVYFVMFVMSAVVPITAMFPNLGYPC
                     YFNALVDYGALNLTNYNLAHHLTPTLYLEPPEMFVYITLVFIADCVAFIYYACGEVAL
                     IKARKKVSGLTDLSAWVSAVGSPTVLFLAILKLWSIQVFIQVLSYKHVFLSAFVYFLH
                     FLASVLHACACVTRFSPVWVVKAQDNSIPQDTFLWWVVFYLKPVVTNLYLGCLALETL
                     VFSLSVFLALGNSFYFMVGDMVLGAVNLFLILPIFWYILTEVWLASFLRHNFGFYCGM
                     FIASIILILPLVRYEAVFVSAKLHTTVAINVAIIPILCSVAMLIRICRIFKSMRQGTD
                     YVPVSETVELELESEPRPRPSRTPSPGRNRRRSSTSSSSSRSTRRQRPVSTQALVSSV
                     LPMTTDSEEEIFP"
     misc_feature    120260
                     /note="ACCEPT: ATCTTCCTCCAGGT possible acceptor"
     misc_feature    120358
                     /note="polyA signal: AATAAA"
     CDS             complement(120747..120974)
                     /note="BBLF1 late reading frame, possibly homologous to RF
                     49 VZV;  NCBI gi: 59131"
```

```
                /codon_start=1
                /translation="MGALWSLCRRRVNSIGDVDGGIINLYNDYEEFNLETTKLIAAEE
                GRACGETNEGLEYDEDSENDELLFLPNKKPN"
misc_feature    complement(120764)
                /note="polyA signal: AATAAA, 3' end of 0.6kb late, 1.6kb
                early, 3.0kb early RNAs"
CDS             complement(120929..122341)
                /note="BGLF5 early reading frame, homologous to RF 48 VZV
                and alkaline exonuclease of HSV;  NCBI gi: 59132"
                /codon_start=1
                /translation="MADVDELEDPMEEMTSYTFARFLRSPETEAFVRNLDRPPQMPAM
                RFVYLYCLCKQIQEFSGETGFCDFVSSLVQENDSKDGPSLKSIYWGLQEATDEQRTVL
                CSYVESMTRGQSENLMWDILRNGIISSSKLLSTIKNGPTKVFEPAPISTNHYFGGPVA
                FGLRCEDTVKDIVCKLICGDASANRQFGFMISPTDGIFGVSLDLCVNVESQGDFILFT
                DRSCIYEIKCRFKYLFSKSEFDPIYPSYTALYKRPCKRSFIRFINSIARPTVEYVPDG
                RLPSEGDYLLTQDEAWNLKDVRKRKLGPGHDLVADSLAANRGVESMLYVMTDPSENAG
                RIGIKDRVPVNIFINPRHNYFYQVLLQYKIVGDYVRHSGGGKPGRDCSPRVNIVTAFF
                RKRSPLDPATCTLGSDLLLDASVEIPVAVLVTPVVLPDSVIRKTLSTAAGSWKAYADN
                TFDTAPWVPSGLFADDESTP"
promoter        complement(121331)
                /note="TATA: TATTAAA BBL1 late promoter before BBLF1"
promoter        121669
                /note="TATA: CATAAAT"
promoter        121697
                /note="TATA: TATAAAG"
promoter        121772
                /note="TATA: CATAAAG"
misc_feature    122313
                /note="BAM: Bam H1 B/G"
CDS             complement(122325..123692)
                /note="BGLF4 early reading frame, homologous to RF 47 VZV;
                 NCBI gi: 583897"
                /codon_start=1
                /translation="XGWRSSVSRSLRPETSCDRPSSHLRNMDVNMAAELSPTNSSSSG
                ELSVSPEPPRETQAFLGKVTVIDYFTFQHKHLKVTNIDDMTETLYVKLPENMTRCDHL
                PITCEYLLGRGSYGAVYAHADNATVKLYDSVTELYHELMVCDMIQIGKATAEDGQDKA
                LVDYLSACTSCHALFMPQFRCSLQDYGHWHDGSIEPLVRGFQGLKDAVYFLNRHCGLF
                HSDISPSNILVDFTDTMWGMGRLVLTDYGTASLHDRNKMLDVRLKSSKGRQLYRLYCQ
                REPFSIAKDTYKPLCLLSKCYILRGAGHIPDPSACGPVGAQTALRLDLQSLGYSLLYG
                IMHLADSTHKIPYPNPDMGFDRSDPLYFLQFAAPKVVLLEVLSQMWNLNLDMGLTSCG
                ESPCVDVTAEHMSQFLQWCRSLKKRFKESYFFNCRPRFEHPHLPGLVAELLADDFFGP
                DGRRG"
misc_feature    complement(123506)
                /note="DONOR: AAGGTGACT possible donor"
CDS             complement(123941..124939)
                /note="BGLF3 reading frame;  NCBI gi: 59135"
                /codon_start=1
                /translation="MFNAVKADMPDDPMLARRYGQCLELALEACQDTPEQFKLVETPL
                KSFLLVSNILPQDNRPWHEARSSGRVAEDDYDFSSLALELLPLNPRLPEEWQFGGQGW
                SSRMEPSQPEMGMGLCFEVFDGDLMRIALAWNKDEVIGQALQILAHSQTWTSLVPEDP
                LPWMWALFYGPRSHCEERHCVYAAARGKRGPILLPTAVYTPCANIEAFLAHLTRCVYA
                LYLDVRDWKGEDIAPPFDVSRLNKMAKQLCLLPQEPFCITRVCLLCLLHKQNLNAQYK
                RPVDTYDPCLILTGEAERYMVDAVGNYREASTGTTVLYPTYDLGSIVADMVTYEDE"
promoter        complement(124117)
                /note="TATA: TATAAAA"
misc_feature    complement(124219)
                /note="polyA signal: AATAAA"
CDS             124938..125915
                /note="BGRF1 reading frame, homologous to RF 45 VZV and
                spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
```



```
                Northern blots in BGRF1 detect 2.7, 2.6, 2.1kb late and
                1.9kb early RNAs. 2.6, 2.1kb RNAs very weak.  NCBI gi:
                59134"
                /codon_start=1
                /translation="MLYASQRGRLTENLRNALQQDSTTQGCLGAETPSIMYTGAKSDR
                WAHPLVGTIHASNLYCPMLRAYCRHYGPRPVFVASDESLPMFGASPALHTPVQVQMCL
                LPELRDTLQRLLPPPNLEDSEALTEFKTSVSSARAILEDPNFLEMREFVTSLASFLSG
                QYKHKPARLEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMTLEKLHIFK
                QKASVFLIPRRHGKTWIVVAIISLILSNLSNVQIGYVAHQKHVASAVFTEIIDTLTKS
                FDSKRVEVNKETSTITFRHSGKISSTVMCATCFNKNVRPDVSVLGNCRA"
promoter        complement(125113)
                /note="TATA: TATAAAT before BGLF3"
misc_feature    complement(125484)
                /note="polyA signal: AATAAA, 3' end of 1.6kb late, 1.8kb
                late, 3.0kb late and 3.7kb early RNAs"
CDS             complement(125863..126873)
                /note="BGLF2 late reading frame, poor homology to RF 44
                VZV;  NCBI gi: 59136"
                /codon_start=1
                /translation="MASAANSSREQLRKFLNKECLWVLSDASTPQMKVYTATTAVSAV
                YVPQIAGPPKTYMNVTLIVLKPKKKPTYVTVYINGTLATVARPEVLFTKAVQGPHSLT
                LMYFGVFSDAVGEAVPVEIRGNPVVTCTDLTTAHVFTTSTAVKTVEELQDITPSEIIP
                LGRGGAWYAEGALYMFFVNMDMLMCCPNMPTFPSLTHFINLLTRCDNGECVTCYGAGA
                HVNILRGWTEDDSPGTSGTCPCLLPCTALNNDYVPITGHRALLGLMFKPEDAPFVVGL
                RFNPPKMHPDMSRVLQGVLANGKEVPCTAQPWTLLRFSDLYSRAMLYNCQVLKRQVLH
                SY"
promoter        126277
                /note="TATA: GATAAAA"
CDS             complement(126851..128374)
                /note="BGLF1 late reading frame;  NCBI gi: 59137"
                /codon_start=1
                /translation="MDVHIDNQVLSGLGTPLLVHLFVPDTVMAELCPNRVPNCEGAWC
                QTLFSDRTGLTRVCRVFAARGMLPGRPSHRGTFTSVPVYCDEGLPELYNPFHVAALRF
                YDEGGLVGELQIYYLSLFEGAKRALTDGHLIREASGVQESAAAMQPIPIDPGPPGGAG
                IEHMPVAAAQVEHPKTYDLKQILLEITQEENRGEQRLGHAGSPALCLGLRLRAGAETK
                AAAETSVSKHHPALENPSNIRGSAGGEGGGGRAGTGGTVGVGSGALSRVPVSFSKTRR
                AIRESRALVRGIAHIFSPHALYVVTYPELSAQGRLHRMTAVTHASPATDLAEVSILGA
                PEREFRFLISVALRISASFREKLAMQAWTAQQEIPVVIPTSYSRIYKNSDLIREAFFT
                VQTRVSWESCWVKATISNAPKTPDACLWIDSHPLYEEGASAWGKVIDSRPPGGLVGAA
                SQLVALGTDGHCVHLATTSDGQAFLVLPGGFVIKGQLALTPEERGYILARHGIRREQ"
promoter        complement(126929)
                /note="TATA: TATTAAA EEL8 late promoter before BGLF2,
                gives 1.6kb late RNA"
promoter        complement(127237)
                /note="TATA: TATAAAA, potential promoter for 1.8kb late
                RNA"
misc_feature    128029
                /note="polyA signal: AATAAA"
CDS             complement(128344..129021)
                /note="BDLF4 early reading frame;  NCBI gi: 59138"
                /codon_start=1
                /translation="MSDQGRLSLPRGEGGTDEPNPRHLCSYSKLEFHLPLPESMASVF
                ACWGCGEYHVCDGSSECTLIETHEGVVCALTGNYMGPHFQPALRPWTEIRQDTQDQRD
                KWEPEQVQGLVKTVVNHLYHYFLNENVISGVSEALFDQEGALRPHIPALVSFVFPCCL
                MLFRGASSEKVVDVVLSLYIHVIISIYSQKTVYGALLFKSTRNKRYDAVAKRMRELWM
                STLTTKC"
promoter        complement(128432)
                /note="TATA: TATTTAA before BGLF1, potential promoter for
                3.0kb late RNA"
misc_feature    128848
```

```
                /note="BAM: Bam H1 G/D"
promoter        complement(129054)
                /note="TATA: TATTTGC before BDLF4, potential promoter for
                3.7kb early RNA"
CDS             129188..130351
                /note="BDRF1 reading frame, homologous to RF 42 VZV and
                spliced gene in HSV (Costa et al, 1985). Spliced from
                BGRF1. Northern blots in BDRF1 detect 2.7, 2.6 kb late and
                1.9kb early RNAs. Possibly also 1.8kb early RNA.  NCBI gi:
                583898"
                /codon_start=1
                /translation="XLASYPVFQSIRGQTFHLLFVDEANFIKKEALPAILGFMLQKDA
                KIIFISSVNSADQATSFLYKLKDAQERLLNVVSYVCQEHRQDFDMQDSMVSCPCFRLH
                IPSYITMDSNIRATTNLFLDGAFSTELMGDTSSLSQGSLSRTVRDDAINQLELCRVDT
                LNPRVAGRLASSLYVYVDPAYTNNTSASGTGIAAVTHDRADPNRVIVLGLEHFFLKDL
                TGDAALQIATCVVALVSSIVTLHPHLEEVKVAVEGNSSQDSAVAIASIIGESCPLPCA
                FVHTKDKTSSLQWPMYLLTNEKSKAFERLIYAVNTASLSASQVTVSNTIQLSFDPVLY
                LISQIRAIKPIPLRDGTYTYTGKQRNLSDDVLVALVMAHFLATTQKHTFKKVH"
promoter        129374
                /note="TATA: TATAAGC"
promoter        complement(129377)
                /note="TATA: TATAAAG"
misc_feature    129413
                /note="DONOR: GTGGTAAGT possible donor"
misc_feature    130347
                /note="polyA signal: ATTAAA"
misc_feature    complement(130359)
                /note="polyA signal: AATAAA, 3' end of 0.9kb late RNA,
                2.3kb late RNA and 3.2kb late RNA"
CDS             complement(130362..131066)
                /note="BDLF3 late reading frame 9xNXT/S;  NCBI gi: 59140"
                /codon_start=1
                /translation="MAHARDKAGAVMAMILICETSLIWTSSGSSTASAGNVTGTTAVT
                TPSPSASGPSTNQSTTLTTTSAPITTTAILSTNTTTVTFTGTTVTPVPTTSNASTINV
                TTKVTAQNITATEAGTGTSTGVTSNVTTRSSSTTSATTRITNATTLAPTLSSKGTSNA
                TKTTAELPTVPDERQPSLSYGLPLWTLVFVGLTFLMLILIFAAGLMMSAKNKPLDEAL
                LTNAVTRDPSLYKGLV"
promoter        complement(131104)
                /note="TATA: TATAAAA EEL4 late promoter before BDLF3,
                gives 0.9kb late RNA"
CDS             complement(131127..132389)
                /note="BDLF2 late reading frame;  NCBI gi: 59141"
                /codon_start=1
                /translation="MVDEQVAVEHGTVSHTISREEDGVVHERRVLASGERVEVFYKAP
                APRPREGRASTFHDFTVPAAAAVPGPEPEPEPHPPMPIHANGGGETKTNTQDQNQNQT
                TRTRTNAKAEERTAEMDDTMASSGGQRGAPISADLLSLSSLTGRMAAMAPSWMKSEVC
                GERMRFKEDVYDGEAETLAEPPRCFMLSFVFIYYCCYLAFLALLAFGFNPLFLPSFMP
                VGAKVLRGKGRDFGVPLSYGCPTNPFCKVYTLIPAVVINNVTYYPNNTDSHGGHGGFE
                AAALHVAALFESGCPNLQAVTNRNRTFNVTRASGRVERRLVQDMQRVLASAVVVMHHH
                CHYETYYVFDGVGPEFGTIPTPCFKDVLAFRPSLVTNCTAPLKTSVKGPNWSGAAGGM
                KRKQCRVDRLTDRSFPAYLEEVMYVMVQ"
promoter        132266
                /note="TATA: TATAAAA"
CDS             complement(132400..133305)
                /note="BDLF1 late reading frame, poor homology to RF 41
                VZV;  NCBI gi: 59142"
                /codon_start=1
                /translation="MDLKVVVSLSSRLYTDEIAKMQQRIGCILPLASTHGTQNVQGLG
                LGQVYSLETVPDYVSMYNYLSDCTLAVLDEVSVDSLILTKIVPGQTYAIKNKYQPFFQ
                WHGTGSLSVMPPVFGREHATVKLESNDVDIVFPMVLPTPIAEEVLQKILLFNVYSRVV
```



```
                MQAPGNADMLDVHMHLGSVSYLGHHYELALPEVPGPLGLALLDNLSLYFCIMVTLLPR
                ASMRLVRGLIRHEHHDLLNLFQEMVPDEIARIDLDDLSVADDLSRMRVMMTYLQSLAS
                LFNLGPRLATAAYSQETLTATCWLR"
promoter        complement(132476)
                /note="TATA: TATTTAA before BDLF2, likely promoter for
                2.3kb late RNA"
misc_feature    complement(133312)
                /note="polyA signal: AATAAA, 3' end of 4.5kb late RNA"
CDS             complement(133321..137466)
                /note="BcLF1 late reading frame, homologous to RF 40 VZV
                and major capsid protein of HSV;  NCBI gi: 59143"
                /codon_start=1
                /translation="MASNEGVENRPFPYLTVDADLLSNLRQSAAEGLFHSFDLLVGKD
                AREAGIKFEVLLGVYTNAIQYVRFLETALAVSCVNTEFKDLSRMTDGKIQFRISVPTI
                AHGDGRRPSKQRTFIVVKNCHKHHISTEMELSMLDLEILHSIPETPVEYAEYVGAVKT
                VASALQFGVDALERGLINTVLSVKLRHAPPMFILQTLADPTFTERGFSKTVKSDLIAM
                FKRHLLEHSFFLDRAENMGSGFSQYVRSRLSEMVAAVSGESVLKGVSTYTTAKGGEPV
                GGVFIVTDNVLRQLLTFLGEEADNQIMGPSSYASFVVRGENLVTAVSYGRVMRTFEHF
                MARIVDSPEKAGSTKSDLPAVAAGVEDQPRVPISAAVIKLGNHAVAVESLQKMYNDTQ
                SPYPLNRRMQYSYYFPVGLFMPNPKYTTSAAIKMLDNPTQQLPVEAWIVNKNNLLLAF
                NLQNALKVLCHPRLHTPAHTLNSLNAAPAPRDRRETYSLQHRRPNHMNVLVIVDEFYD
                NKYAAPVTDIALKCGLPTEDFLHPSNYDLLRLELHPLYDIYIGRDAGERARHRAVHRL
                MVGNLPTPLAPAAFQEARGQQFETATSLAHVVDQAVIETVQDTAYDTAYPAFFYVVEA
                MIHGFEEKFVMNVPLVSLCINTYWERSGRLAFVNSFSMIKFICRHLGNNAISKEAYSM
                YRKIYGELIALEQALMRLAGSDVVGDESVGQYVCALLDPNLLPPVAYTDIFTHLLTVS
                DRAPQIIIGNEVYADTLAAPQFIERVGNMDEMAAQFVALYGYRVNGDHDHDFRLHLGP
                YVDEGHADVLEKIFYYVFLPTCTNAHMCGLGVDFQHVAQTLAYNGPAFSHHFTRDEDI
                LDNLENGTLRDLLEISDLRPTVGMIRDLSASFMTCPTFTRAVRVSVDNDVTQQLAPNP
                ADKRTEQTVLVNGLVAFAFSERTRAVTQCLFHAIPFHMFYGDPRVAATMHQDVATFVM
                RNPQQRAVEAFNRPEQLFAEYREWHRSPMGKYAAECLPSLVSISGMTAMHIKMSPMAY
                IAQAKLKIHPGVAMTVVRTDEILSENILFSSRASTSMFIGTPNVSRREARVDAVTFEV
                HHEMASIDTGLSYSSTMTPARVAAITTDMGIHTQDFFSVFPAEAFGNQQVNDYIKAKV
                GAQRNGTLLRDPRTYLAGMTNVNGAPGLCHGQQATCEIIVTPVTADVAYFQKSNSPRG
                RAACVVSCENYNQEVAEGLIYDHSRPDAAYEYRSTVNPWASQLGSLGDIMYNSSYRQT
                AVPGLYSPCRAFFNKEELLRNNRGLYNMVNEYSQRLGGHPATSNTEVQFVVIAGTDVF
                LEQPCSFLQEAFPALSASSRALIDEFMSVKQTHAPIHYGHYIIEEVAPVRRILKFGNK
                VVF"
misc_feature    complement(133332)
                /note="DONOR: AAGGTGGTT possible donor"
promoter        complement(133352)
                /note="TATA: TATTAAA before BDLF1"
promoter        complement(133386)
                /note="TATA: TATATAA"
misc_feature    135178
                /note="polyA signal: AATAAA"
promoter        135394
                /note="TATA: TATAAGT"
misc_feature    136624
                /note="polyA signal: AATAAA"
misc_feature    136868
                /note="BAM: Bam H1 D/c"
promoter        complement(137710)
                /note="TATA: TATTAAA EHL1 promoter before BcLF1, gives
                4.5kb late RNA"
promoter        137857
                /note="TATA: CATAAAC"
CDS             137862..139718
                /note="BcRF1 reading frame;  NCBI gi: 583899"
                /codon_start=1
                /translation="XRAQCFALSSATCLITNLSAASSSSAAWSRVGFLGTCKRNSAKM
```

```
                LAHLNQVTRIPPCPPFSGREARLKFHFFSWSTFMLSWPNNATLREIRTRAATNLTHHP
                HLVDTLYHASPQTPFLTRSGALYRFVTCCNCTLPNISIQQCKAGDRPGDLEIILQSNG
                GGRPASFQFPSSPTGSLLRCIVAASLLPEVSVGHQELSPLRSRSQGGQTDVRSGPDPA
                RRLVALLRREDGAPKDPPLGPFGHPRGPGPAKSEDEESERRDAPPPPLDSSFQASRLV
                PVGPGFRLLVFNTNRVINTKLVCSEPLVKMRVCNVPRLINNFVARKYVVKETAFTVSL
                FFTDGVGANLAINVNISGTYLSFLLAMTSLRCFLPVEAIYPAAVSNWNSTLDLHGLEN
                QSLVRENRSGVFWTTNFPSVVSCRDGLNVSWFKAATATISRVHGQTLEQHLIREITPI
                VTHREAKISRIKNRLFTLLELRNRSQIQVLHKRFLEGLLDCASLLRLDPSCINRIASE
                GLFDFSKRSIAHSKNRHECALLGHRHSANVTKLVVNERKTRLDILGRNANFLTRCKHQ
                VNLRQSPIFLTLLRHIRRRLGLGRASVKREITLLLAHLRKKTAPIHCRDAQV"
misc_feature    138019
                /note="BAM: Bam H1 c/b"
misc_feature    139352
                /note="BAM: Bam H1 b/T"
CDS             139642..140919
                /note="BTRF1 reading frame. Northern blots detect 0.95
                late and 3.8kb early RNA;  NCBI gi: 583900"
                /codon_start=1
                /translation="XERLPFSWPTCAKRQPPSTAVMLKCKQPGARFIHGAVHLPSGQI
                VFHTIHSPTLASALGLPGENVPIPALFRASGLNVRESLPMTNMRAPIISLARLILAPN
                PYILEGQLTVGMTQDNGIPVLFARPVIEVKSGPESNIKASSQLMIAEDSCLNQIAPFS
                ASEHPAFSMVESVKRVRVDEGANTRRTIRDILEIPVTVLSSLQLSPTKSILKKAPEPP
                PPEPQATFDATPYARIFYDIGRQVPKLGNAPAAQVSNVLIANRSHNSLRLVPNPDLLP
                LQHLYLKHVVLKSLNLENIVQDFEAIFTSPSDTISEAETKAFEKLVEQAKNTVENIVF
                CLNSICSTSTLPDVVPDVNNPNISLALEKYFLMFPPSGTIMRNVRFATPIVRLLCQGA
                ELGTMAQFLGKYIKVKKETGMYTLVKLYYLLRI"
misc_feature    complement(140902)
                /note="polyA signal: AATAAA, 3' end of 2.5kb late RNA"
CDS             complement(140916..143036)
                /note="BXLF2 late reading frame, encodes gp85; homologous
                to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV);
                NCBI gi: 59146"
                /codon_start=1
                /translation="MQLLCVFCLVLLWEVGAASLSEVKLHLDIEGHASHYTIPWTELM
                AKVPGLSPEALWREANVTEDLASMLNRYKLIYKTSGTLGIALAEPVDIPAVSEGSMQV
                DASKVHPGVISGLNSPACMLSAPLEKQLFYYIGTMLPNTRPHSYVFYQLRCHLSYVAL
                SINGDKFQYTGAMTSKFLMGTYKRVTEKGDEHVLSLVFGKTKDLPDLRGPFSYPSLTS
                AQSGDYSLVIVTTFVHYANFHNYFVPNLKDMFSRAVTMTAASYARYVLQKLVLLEMKG
                GCREPELDTETLTTMFEVSVAFFKVGHAVGETGNGCVDLRWLAKSFFELTVLKDIIGI
                CYGATVKGMQSYGLERLAAMLMATVKMEELGHLTTEKQEYALRLATVGYPKAGVYSGL
                IGGATSVLLSAYNRHPLFQPLHTVMRETLFIGSHVVLRELRLNVTTQGPNLALYQLLS
                TALCSALEIGEVLRGLALGTESGLFSPCYLSLRFDLTRDKLLSMAPQEATLDQAAVSN
                AVDGFLGRLSLEREDRDAWHLPAYKCVDRLDKVLMIIPLINVTFIISSDREVRGSALY
                EASTTYLSSSLFLSPVIMNKCSQGAVAGEPRQIPKIQNFTRTQKSCIFCGFALLSYDE
                KEGLETTTYITSQEVQNSILSSNYFDFDNLHVHYLLLTTNGTVMEIAGLYEERAHVVL
                AIILYFIAFALGIFLVHKIVMFFL"
misc_feature    140970
                /note="polyA signal: AATAAA"
misc_feature    complement(141286)
                /note="polyA signal: AATAAA"
promoter        142589
                /note="TATA: GATAAAA"
misc_feature    142740
                /note="BAM: Bam H1 T/X"
CDS             complement(143038..144861)
                /note="BXLF1 early reading frame, thymidine kinase
                (Littler et al, 1986). Weak homology to RF 36 VZV and HSV
                thymidine kinase. 4.0kb early RNA presumably encodes the
                TK. Also a 2.2kb late RNA here.  NCBI gi: 59148"
                /codon_start=1
```



```
                /translation="MAGFPGKEAGPPGGWRKCQEDESPENERHENFYAEIDDFAPSVL
                TPTGSDSGAGEEDDDGLYQVPTHWPPLMAPTGLSGERVPCRTQAAVTSNTGNSPGSRH
                TSCPFTLPRGAQPPAPAHQKPTAPTPKPRSRECGPSKTPDPFSWFRKTSCTEGGADST
                SRSFMYQKGFEEGLAGLGLDDKSDCESEDESNFRRPSSHSALKQKNGGKGKPSGLFEH
                LAAHGREFSKLSKHAAQLKRLSGSVMNVLNLDDAQDTRQAKAQRKESMRVPIVTHLTN
                HVPVIKPACSLFLEGAPGVGKTTMLNHLKAVFGDLTIVVPEPMRYWTHVYENAIKAMH
                KNVTRARHGREDTSAEVLACQMKFTTPFRVLASRKRSLLVTESGARSVAPLDCWILHD
                RHLLSASVVFPIMLLRSQLLSYSDFIQVLATFTADPGDTIVWMKLNVEENMRRLKKRG
                RKHESGLDAGYLKSVNDAYHAVYCAWLLTQYFAPEDIVKVCAGLTTITTVCHQSHTPI
                IRSGVAEKLYKNSIFSVLKEVIQPFRADAVLLEVCLAFTRTLAYLQFVLVDLSEFQDD
                LPGCWTEIYMQALKNPAIRSQFFDWAGLSKVISDFERGNRD"
promoter        complement(143310)
                /note="TATA: TATAAGA ECL2 late promoter before BXLF2,
                gives 2.5kb late RNA"
misc_feature    143608
                /note="polyA signal: AATAAA"
misc_feature    144791
                /note="ACCEPT: TCTTTCGTTTTCAGG poss. acceptor before
                BXRF1"
CDS             144860..145606
                /note="BXRF1 late reading frame, homologous to RF 35 VZV.
                Basic (core?) protein.  NCBI gi: 59147"
                /codon_start=1
                /translation="MDPTRGLCALSTHDLAKFHSLPPARKAAGKRAHLRCYSKLLSLK
                SWEQLASFLSLPPGPTFTDFRLFFEVTLGRRIADCVVVALQPYPRCYIVEFKTAMSNT
                ANPQSVTRKAQRLEGTAQLCDCANFLRTSCPPVLGSQGLEVLAALVFKNQRSLRTLQV
                EFPALGQKTLPTSTTGLLNLLSRWQDGALRARLDRPRPTAQGHRPRTHVGPKPSQLTA
                RVPRSARAGRAGGRKGQVGAVGQVCPGAQK"
misc_feature    144862
                /note="BAM: Bam H1 X/V"
misc_feature    144945
                /note="DONOR: CAGGTAAGC possible donor at 3' BXRF1"
promoter        complement(145135)
                /note="TATA: TATAACA before BXLF1"
promoter        145302
                /note="TATA: TATTTAA before BVRF1, potential promoter for
                1.9kb early RNA"
CDS             145416..147128
                /note="BVRF1 early reading frame, homologous to RF 34 VZV;
                 NCBI gi: 59149"
                /codon_start=1
                /translation="MALSGHVLIDPARLPRDTGPELMWAPSLRNSLRVSPEALELAER
                EAERARSERWDRCAQVLKNRLLRVELDGIMRDHLARAEEIRQDLDAVVAFSDGLESMQ
                VRSPSTGGRSAPAPPSPSPAQPFTRLTGNAQYAVSISPTDPPLMVAGSLAQTLLGNLY
                GNINQWVPSFGPWYRTMSANAMQRRVFPKQLRGNLNFTNSVSLKLMTEVVAVLEGTTQ
                DFFSDVRHLPDLQAALILSVAYLLLQGGSSHQQRPLPASREELLELGPESLEKIIADL
                KAKSPGGNFMILTSGNKEARQSIAPLNRQAAYPPGTFADNKIYNLFVGAGLLPTTAAL
                NVPGAAGRDRDLVYRIANQIFGEDVPPFSSHQWNLRVGLAALEALMLVYTLCETANLA
                EAATRRLHLSSLLPQAMQRRKPAMASAGMPGAYPVQTLFRHGELFRFIWAHYVRPTVA
                ADPQASISSLFPGLVLLALELKLMDGQAPSHYAINLTGQKFDTLFEIINQKLLFHDPA
                AMLAARTQLRLAFEDGVGVALGRPSPMLAAREILERQFSASDDYDRLYFLTLGYLASP
                VAPS"
misc_feature    complement(146926)
                /note="polyA signal: AATAAA"
misc_feature    147167
                /note="DONOR: AAGGTAAAT possible donor"
misc_feature    147170
                /note="polyA signal: AATAAA, 3' end of 2.4kb late and
                1.9kb early RNAs"
promoter        147721
```

```
                /note="TATA: TATTTAT before BVRF2, potential promoter for
                2.1kb early RNA"
CDS             147927..149744
                /note="BVRF2 early reading frame, N-terminus homologous to
                RF 33 VZV;  NCBI gi: 59150"
                /codon_start=1
                /translation="MVQAPSVYVCGFVERPDAPPKDACLHLDPLTVKSQLPLKKPLPL
                TVEHLPDAPVGSVFGLYQSRAGLFSAASITSGDFLSLLDSIYHDCDIAQSQRLPLPRE
                PKVEALHAWLPSLSLASLHPDIPQTTADGGKLSFFDHVSICALGRRRGTTAVYGTDLA
                WVLKHFSDLEPSIAAQIENDANAAKRESGCPEDHPLPLTKLIAKAIDAGFLRNRVETL
                RQDRGVANIPAESYLKASDAPDLQKPDKALQSPPPASTDPATMLSGNAGEGATACGGS
                AAAGQDLISVPRNTFMTLLQTNLDNKPPRQTPLPYAAPLPPFSHQAIATAPSYGPGAG
                AVAPAGGYFTSPGGYYAGPAGGDPGAFLAMDAHTYHPHPHPPPAYFGLPGLFGPPPPV
                PPYYGSHLRADYVPAPSRSNKRKRDPEEDEEGGGLFPGEDATLYRKDIAGLSKSVNEL
                QHTLQALRRETLSYGHTGVGYCPQQGPCYTHSGPYGFQPHQSYEVPRYVPHPPPPPTS
                HQAAQAQPPPPGTQAPEAHCVAESTIPEAGAAGNSGPREDTNPQQPTTEGHHRGKKLV
                QASASGVAQSKEPTTPKAKSVSAHLKSIFCEELLNKRVA"
misc_feature    148007
                /note="BAM: Bam H1 V/d"
promoter        148620
                /note="TATA: TATTTAA ECR1 late promoter before BdRF1,
                gives 1.2kb late RNA"
CDS             148707..149744
                /note="BdRF1 reading frame; this is the C terminus of
                BVRF2;  NCBI gi: 59151"
                /codon_start=1
                /translation="MLSGNAGEGATACGGSAAAGQDLISVPRNTFMTLLQTNLDNKPP
                RQTPLPYAAPLPPFSHQAIATAPSYGPGAGAVAPAGGYFTSPGGYYAGPAGGDPGAFL
                AMDAHTYHPHPHPPPAYFGLPGLFGPPPPVPPYYGSHLRADYVPAPSRSNKRKRDPEE
                DEEGGGLFPGEDATLYRKDIAGLSKSVNELQHTLQALRRETLSYGHTGVGYCPQQGPC
                YTHSGPYGFQPHQSYEVPRYVPHPPPPPTSHQAAQAQPPPPGTQAPEAHCVAESTIPE
                AGAAGNSGPREDTNPQQPTTEGHHRGKKLVQASASGVAQSKEPTTPKAKSVSAHLKSI
                FCEELLNKRVA"
misc_feature    149115
                /note="BAM: Bam H1 d/I"
misc_feature    149727
                /note="polyA signal: AATAAA, 3' end of 2.1kb early and
                1.2kb late RNAs"
misc_feature    complement(149758)
                /note="polyA signal: AATAAA, 3' end of 1.0kb late, 1.5kb
                late and 1.8kb late RNAs"
CDS             complement(149779..150525)
                /note="BILF2 late reading frame 11xNXT/S;  NCBI gi: 59152"
                /codon_start=1
                /translation="MTHLVLLLCCCVGSVCAFFSDLVKFENVTAHAGARVNLTCSVPS
                NESVSRIELGRGYTPGDGQLPLAVATSNNGTHITNGGYNYSLTLEWVNDSNTSVSLII
                PNVTLAHAGYYTCNVTLRNCSVASGVHCNYSAGEEDDQYHANRTLTQRMHLTVIPATT
                IAPTTLVSHTTSTSHRPHRRPVSKRPTHKPVTLGPFPIDPWRPKTTWVHWALLLITCA
                VVAPVLLIIIISCLGWLAGWGRRRKGWIPL"
promoter        complement(150571)
                /note="TATA: TATTTAG before BILF2. Potential promoter for
                1.0kb late RNA."
repeat_region   151236..151618
                /note="repetitive sequence 3X"
misc_feature    151767
                /note="polyA signal: AATAAA"
promoter        complement(151780)
                /note="TATA: CATAAAA"
misc_feature    152012..152013
                /note="DEL: B95-8 deletion with respect to Raji"
```



```
CDS             complement(152161..153099)
                /note="BILF1 reading frame, membrane protein, 3xNXS/T;
                NCBI gi: 59153"
                /codon_start=1
                /translation="MLSTMAPGSTVGTLVANMTSVNATEDACTKSYSAFLSGMTSLLL
                VLLILLTLAGILFIIFVRKLVHRMDVWLIALLIELLLWVLGKMIQEFSSTGLCLLTQN
                MMFLGLMCSVWTHLGMALEKTLALFSRTPKRTSHRNVCLYLMGVFCLVLLLIIILLIT
                MGPDANLNRGPNMCREGPTKGMHTAVQGLKAGCYLLAAVLIVLLTVIIIWKLLRTKFG
                RKPRLICNVTFTGLICAFSWFMLSLPLLFLGEAGSLGFDCTESLVARYYPGPAACLAL
                LLIILYAWSFSHFMDSLKNQVTVTARYFRRVPSQST"
promoter        152230
                /note="TATA: CATAAAA"
misc_feature    153259
                /note="polyA signal: AATAAA"
misc_feature    153637
                /note="HPN: 22bp 2-fold symmetric"
misc_feature    complement(153690)
                /note="DONOR: AAAGTGAGG possible donor"
CDS             complement(153699..156746)
                /note="BALF5 DNA polymerase (early), homologous to many
                DNA polymerases, CMV HFLF2 and RF 28 VZV. 4.5kb early RNA
                apparently encodes BALF5, RNA ends unknown.  NCBI gi:
                59154"
                /codon_start=1
                /translation="MSGGLFYNPFLRPNKGLLKKPDKEYLRLIPKCFQTPGAAGVVDV
                RGPQPPLCFYQDSLTVVGGDEDGKGMWWRQRAQEGTARPEADTHGSPLDFHVYDILET
                VYTHEKCAVIPSDKQGYVVPCGIVIKLLGRRKADGASVCVNVFGQQAYFYASAPQGLD
                VEFAVLSALKASTFDRRTPCRVSVEKVTRRSIMGYGNHAGDYHKITLSHPNSVCHVAT
                WLQDKHGCRIFEANVDATRRFVLDNDFVTFGWYSCRRAIPRLQHRDSYAELEYDCEVG
                DLSVRREDSSWPSYQALAFDIECLGEEGFPTATNEADLILQISCVLWSTGEEAGRYRR
                ILLTLGTCEDIEGVEVYEFPSELDMLYAFFQLIRDLSVEIVTGYNVANFDWPYILDRA
                RHIYSINPASLGKIRAGGVCEVRRPHDAGKGFLRANTKVRITGLIPIDMYAVCRDKLS
                LSDYKLDTVARHLLGAKKEDVHYKEIPRLFAAGPEGRRRLGMYCVQDSALVMDLLNHF
                VIHVEVAEIAKIAHIPCRRVLDDGQQIRVFSCLLAAAQKENFILPMPSASDRDGYQGA
                TVIQPLSGFYNSPVLVVDFASLYPSIIQAHNLCYSTMITPGEEHRLAGLRPGEDYESF
                RLTGGVYHFVKKHVHESFLASLLTSWLAKRKAIKKLLAACEDPRQRTILDKQQLAIKC
                TCNAVYGFTGVANGLFPCLSIAETVTLQGRTMLERAKAFVEALSPANLQALAPSPDAW
                APLNPEGQLRVIYGDTDSLFIECRGFSESETLRFADALAAHTTRSLFVAPISLEAEKT
                FSCLMLITKKRYVGVLTDGKTLMKGVELVRKTACKFVQTRCRRVLDLVLADARVKEAA
                SLLSHRPFQESFTQGLPVGFLPVIDILNQAYTDLREGRVPMGELCFSTELSRKLSAYK
                STQMPHLAVYQKFVERNEELPQIHDRIQYVFVEPKGGVKGARKTEMAEDPAYAERHGV
                PVAVDHYFDKLLQGAANILQCLFDNNSGAALSVLQNFTARPPF"
misc_feature    154747
                /note="BAM: Bam H1 I/A"
misc_feature    complement(156707)
                /note="polyA signal: AATAAA; 3' end of 2.5kb late (gB) RNA
                and 1.8kb late RNA"
CDS             complement(156749..159322)
                /note="BALF4 late reading frame 9xNXT/S homologous to HSV1
                glycoprotein B (Pellet et al, 1985), CMV HFLF1 and RF 31
                VZV (gpII);  NCBI gi: 59155"
                /codon_start=1
                /translation="MTRRRVLSVVVLLAALACRLGAQTPEQPAPPATTVQPTATRQQT
                SFPFRVCELSSHGDLFRFSSDIQCPSFGTRENHTEGLLMVFKDNIIPYSFKVRSYTKI
                VTNILIYNGWYADSVTNRHEEKFSVDSYETDQMDTIYQCYNAVKMTKDGLTRVYVDRD
                GVNITVNLKPTGGLANGVRRYASQTELYDAPGWLIWTYRTRTTVNCLITDMMAKSNSP
                FDFFVTTTGQTVEMSPFYDGKNKETFHERADSFHVRTNYKIVDYDNRGTNPQGERRAF
                LDKGTYTLSWKLENRTAYCPLQHWQTFDSTIATETGKSIHFVTDEGTSSFVTNTTVGI
                ELPDAFKCIEEQVNKTMHEKYEAVQDRYTKGQEAITYFITSGGLLLAWLPLTPRSLAT
                VKNLTELTTPTSSPPSSPSPPAPSAARGSTPAAVLRRRRRDAGNATTPVPPTAPGKSL
```

GTLNNPATVQIQFAYDSLRRQINRMLGDLARAWCLEQKRQNMVLRELTKINPTTVMSS
IYGKAVAAKRLGDVISVSQCVPVNQATVTLRKSMRVPGSETMCYSRPLVSFSFINDTK
TYEGQLGTDNEIFLTKKMTEVCQATSQYYFQSGNEIHVYNDYHHFKTIELDGIATLQT
FISLNTSLIENIDFASLELYSRDEQRASNVFDLEGIFREYNFQAQNIAGLRKDLDNAV
SNGRNQFVDGLGELMDSLGSVGQSITNLVSTVGGLFSSLVSGFISFFKNPFGGMLILV
LVAGVVILVISLTRRTRQMSQQPVQMLYPGIDELAQQHASGEGPGINPISKTELQAIM
LALHEQNQEQKRAAQRAAGPSVASRALQAARDRFPGLRRRRYHDPETAAALLGEAETE
F"

promoter 158204
/note="TATA: TATAAAA"

CDS complement(159309..161678)
/note="BALF3 reading frame; NCBI gi: 583901"
/codon_start=1
/translation="XRRGVLIGPLLRPGGQRPRNPGDHCLQRDRVDGGGRSGLSGGSV
GWARGSPGQLYQRGRAGQHRVAGLGRAAAGWRRAGLRGQTQAPSGHRSPRTRGLETPG
AAMSGLLAAAYSQVYALAVELSVCTRLDPRSLDVAAVVRNAGLLAELEAILLPRLRRQ
NDRACSALSLELVHLLENSREASAALLAPGRKGTRVPPLRTPSVAYSVEFYGGHKVDV
SLCLINDIEILMKRINSVFYCMSHTMGLESLERALDLLGRFRGVSPIPDPRLYITSVP
CWRCVGELMVLPNHGNPSTAEGTHVSCNHLAVPVNPEPVSGLFENEVRQAGLGHLLEA
EEKARPGGPEEGAVPGPGRPEAEGATRALDTYNVFSTVPPEVAELSELLYWNSGGHAI
GATGQGEGGGHSRLSALFARERRLALVRGACEEALAGARLTHLFDAVAPGATERLFCG
GVYSSSGDAVEALKADCAAAFTAHPQYRAILQKRNELYTRLNRAMQRLGRGEEEASRE
SPEVPRPAGAREPGPSGALSDALKRKEQYLRQVATEGLAKLQSCLAQQSETLTETLCL
RVWGDVVYWELARMRNHFLYRRAFVSGPWEDRRAGEGAAFENSKYIKTHLFTQTLSSE
HLHALTHSLYTFITGPLAEESGLFPPPSNVALARCCDAAGTLPHQKAFLTSLIWPGIE
PSDWIETSFNSFYSVPGGSLASSQQILCRALREAVLTVSLYNKTWGRSLILRRADAVS
PGQALPPDGLYLTYDSDRPLILLYKGRGWVFKDLYALLYLHLQMRDDSA"

promoter complement(159370)
/note="TATA: TATTTAA ECL1 late promoter before BALF4, gives 2.5kb late RNA"

misc_feature 160966
/note="polyA signal: AATAAA"

misc_feature 160990
/note="polyA signal: ATTAAA"

misc_feature complement(161013)
/note="polyA signal: AATAAA, presumed end of 3.9kb early RNA"

CDS complement(161384..164770)
/note="BALF2 early reading frame, homologous to RF 29 VZV and major DNA binding protein HSV. 3.9kb early RNA; NCBI gi: 59157"
/codon_start=1
/translation="MQGAQTSEDNLGSQSQPGPCGYIYFYPLATYPLREVATLGTGYA
GHRCLTVPLLCGITVEPGFSINVKALHRRPDPNCGLLRATSYHRDIYVFHNAHMVPPI
FEGPGLEALCGETREVFGYDAYSALPRESSKPGDFFPEGLDPSAYLGAVAITEAFKER
LYSGNLVAIPSLKQEVAVGQSASVRVPLYLDKEVFPEGVPQLRQFYNSDLSRCMHEALY
TGLAQALRVRRVGKLVELLEKQSLQDQAKVAKVAPLKEFPASTISHPDSGALMIVDSA
ACELAVSYAPAMLEASHETPASLNYDSWPLFADCEGPEARVAALHRYNASLAPHVSTQ
IFATNSVLYVSGVSKSTGQGKESLFNSFYMTHGLGTLQEGTWDPCRRPCFSGWGGPDV
TGTNGPGNYAVEHLVYAASFSPNLLARYAYYLQFCQGQKSSLTPVPETGSYVAGAAAS
PMCSLCEGRAPAVCLNTLFFRLRDRFPPVMSTQRRDPYVISGASGSYNETDFLGNFLN
FIDKEDDGQRPDDEPRYTYWQLNQNLLERLSRLGIDAEGKLEKEPHGPRDFVKMFKDV
DAAVDAEVVQFMNSMAKNNITYKDLVKSCYHVMQYSCNPFAQPACPIFTQLFYRSLLT
ILQDISLPICMCYENDNPGLGQSPPEWLKGHYQTLCTNFRSLAIDKGVLTAKEAKVVH
GEPTCDLPDLDAALQGRVYGRRLPVRMSKVLMLCPRNIKIKNRVVFTGENAALQNSFI
KSTTRRENYIINGPYMKFLNTYHKTLFPDTKLSSLYLWHNFSRRRSVPVPSGASAEEY
SDLALFVDGGSRAHEESNVIDVVPGNLVTYAKQRLNNAILKACGQTQFYISLIQGLVP
RTQSVPARDYPHVLGTRAVESAAAYAEATSSLTATTVVCAATDCLSQVCKARPVVTLP
VTINKYTGVNGNNQIFQAGNLGYFMGRGVDRNLLQAPGAGLRKQAGGSSMRKKFVFAT
PTLGLTVKRRTQAATTYEIENIRAGLEAIISQKQEEDCVFDVVCNLVDAMGEACASLT

RDDAEYLLGRFSVLADSVLETLATIASSGIEWTAEAARDFLEGVWGGPGAAQDNFISV
AEPVSTASQASAGLLLGGGGQGSGGRRKRRLATVLPGLEV"

misc_feature 163978..166635
/note="DEL: deletion in Raji"

promoter complement(164814)
/note="TATA: CATTTAA before BALF2, presumed promoter for 3.9kb early RNA"

misc_feature complement(164851)
/note="polyA signal: AATAAA"

CDS complement(164855..165517)
/note="BALF1 early reading frame, 0.7kb early RNA; NCBI gi: 59159"
/codon_start=1
/translation="MNLAIALDSPHPGLASYTILPRPFYHISLKPVSWPDETMRPAKS
TDSVFVRTPVEAWVAPSPPDDKVAESSYLMFRAMYAVFTRDEKDLPLPALVLCRLIKA
SLRKDRKLYAELACRTADIGGKDTHVRLIISVLRAVYNDHYDYWSRLRVVLCYTVVFA
VRNYLDDHKSAAFVLGAIAHYLALYRRLWFARLGGMPRSLRRQFPVTWALASLTDFLK
SL"

promoter 165442
/note="TATA: GATAAAA"

promoter 165466
/note="TATA: TATAAGA EDR1 early promoter before BARF1, gives 0.8kb early RNA"

CDS 165504..166169
/note="BARF1 reading frame; NCBI gi: 59158"
/codon_start=1
/translation="MARFIAQLLLLASCVAAGQAVTAFLGERVTLTSYWRRVSLGPEI
EVSWFKLGPGEEQVLIGRMHHDVIFIEWPFRGFFDIHRSANTFFLVVTAANISHDGNY
LCRMKLGETEVTKQEHLSVVKPLTLSVHSERSQFPDFSVLTVTCTVNAFPHPHVQWLM
PEGVEPAPTAANGGVMKEKDGSLSVAVDLSLPKPWHLPVTCVGKNDKEEAHGVYVSGY
LSQ"

promoter complement(165713)
/note="TATA: TATAAAG before BALF1"

misc_feature 166165
/note="polyA signal: AATAAA 3' end of 0.8kb early RNA. Also 1kb late RNA in this region."

promoter 166469..166475
/note="TATA: TTATTTT"

mRNA 166498..166916
/note="Exon 1 of terminal protein RNA"

misc_feature 166561..166563
/note="Likely initiator met of terminal protein"

misc_feature 166614
/note="BAM: Bam H1 A/Nhet"

misc_feature complement(166946)
/note="polyA signal: AATAAA These 2 polyA sites are 3' end of 0.8kb early, 2.5kb late and 2.5kb latent RNAs"

misc_feature complement(166950)
/note="polyA signal: AATAAA"

CDS complement(166998..167303)
/note="BNLF2b reading frame; NCBI gi: 583902"
/codon_start=1
/translation="XATMRPGRPLAGFYATLRRSFRRMSKRSKNKAKKERVPVEDRPP
TPMPTSQRLIRRNALGGGVRPDAEDCIQRFHPLEPALGVSTKNFDLLSLRCELGWCG"

CDS complement(167304..167486)
/note="BNLF2a reading frame; NCBI gi: 59161"
/codon_start=1
/translation="MVHVLERALLEQQSSACGLPGSSTETRPSHPCPEDPDVSRLRLL
LVVLCVLFGLLCLLLI"

promoter 167320

```
                     /note="TATA: CATAAAA"
     promoter        complement(167525)
                     /note="TATA: TATAAAA EDL2 early promoter before BNLF2a,b.
                     Gives 0.8kb RNA"
     mRNA            complement(168163..168965)
                     /note="BNLF1 coding part of exon c of latent membrane
                     protein"
     repeat_region   168399..168574
                     /note="5 x 33bp repeats"
     intron          complement(168966..169041)
                     /note="intervening sequence in BNLF1"
     misc_feature    complement(168966)
                     /note="ACCEPT: TTTTTTCAACGCAGT possible acceptor"
     misc_feature    complement(169041)
                     /note="DONOR: TTGGTAAGA donor"
     mRNA            complement(169042..169128)
                     /note="BNLF1 exon b of latent membrane protein mRNA"
     intron          complement(169129..169206)
                     /note="intervening sequence in BNLF1"
     misc_feature    complement(169129)
                     /note="ACCEPT: TCCTCTTTCCCCAGT acceptor"
     promoter        complement(169201)
                     /note="TATA: TATTACA EDL1A late promoter, gives 2.5kb late
                     RNA"
     misc_feature    complement(169206)
                     /note="DONOR: TGAGTAAGT donor"
     mRNA            complement(169207..169474)
                     /note="BNLF1 exon a of latent membrane protein mRNA"
     promoter        complement(169546)
                     /note="TATA: TACATAAGC EDL1 promoter before BNLF1 gives
                     2.5kb latent RNA (LMP)"
     repeat_region   170094..170631
                     /note="terminal repeat 1 538bp"
     repeat_region   170632..171154
                     /note="terminal repeat 2 523bp"
     repeat_region   171155..171692
                     /note="terminal repeat 3 538bp"
     repeat_region   171693..172231
                     /note="terminal repeat 4 538bp"
 BASE COUNT    34054 a  52511 c  50755 g  34961 t
 ORIGIN

Initial Score    =    170  Optimized Score  =    454  Significance = 10.52
Residue Identity =    50%  Matches          =    529  Mismatches   =   418
Gaps             =    105  Conservative Substitutions               =     0

             X       10          20        30        40        50
             CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
                |||  |||   ||    |||  ||  | |    ||    |  ||   ||    || |  |     |
    CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
107030    107040    107050    107060    107070    107080    107090    107100

    60        70          80        90          100       110       120
    CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
     |    | |||||  | | ||  |||  ||||  ||  |    |||  | |   | || ||       |  |||||
    GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
        107110    107120    107130    107140    107150    107160    107170

      130       140       150       160       170       180       190
    CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
```



```
    | |   | ||  ||      |     | |  | | ||  || |  ||   ||  |  || || | | | | ||
    CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
          107180    107190    107200      107210    107220    107230

       200       210       220          230       240       250       260
    -TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
      ||  |   |     |   | |    |||  |    |||||| || || |  |  ||||   ||  ||
    GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
107240    107250      107260    107270    107280    107290    107300

          270         280          290       300       310       320
    CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
     |   |||| |||  | |  ||| | ||     ||| | ||||||   |     ||| | || | | |    |
    GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
107310    107320    107330    107340    107350          107360    107370

     330       340       350       360       370          380       390
    CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
      ||        ||  | ||||| |  |||  || | |||||| ||||||  |   | |||||| ||||| ||
    GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
     107380    107390    107400      107410    107420    107430    107440

      400       410       420       430       440       450       460
    GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
    |||| ||| ||||||   | |||||  |  |  ||  | ||    |||| || || |||    || | ||||
    GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
      107450    107460    107470    107480      107490    107500    107510

          470         480       490       500       510       520
    GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
     | ||  ||  ||||    ||  ||| | |     ||||  ||||    ||| | |  | || | | |  |
    AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
      107520    107530    107540    107550    107560    107570    107580

   530       540       550       560       570       580       590       600
    CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
     ||||   || | ||  ||| |   |  || |    |||||||| |||  | |  || | |||  |    |
    AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
    107590    107600      107610    107620    107630    107640    107650

          610         620       630       640       650       660
    AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
    || |||  ||||  || ||  ||  ||||     ||   || | ||  | | |     |     | | ||
    AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
      107660    107670    107680    107690      107700    107710    107720

     670       680       690       700       710       720       730
    AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
    |||| |    |||   || |   |||  ||  | |  |||  ||   | |  | |  | | |||||    |
    AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
      107730    107740    107750    107760    107770    107780    107790

         740       750       760       770       780          790
    AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA----CTGGGTAATCTT
    ||  ||  ||||     ||| |||   |  |  | | |||  | |||  | ||  |    | ||     |
    AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
    107800    107810    107820    107830      107840    107850    107860

  800       810       820       830       840       850       860
    GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
```

```
     GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
         107870    107880    107890    107900    107910    107920

   870       880       890       900       910       920       930
    AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
    AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
07930    107940    107950    107960    107970    107980    107990

  940       950       960           970       980       X
    GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
    GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
 108000    108010    108020    108030    108040    108050


8. US-08-162-407-5' (1-988)
   HS4B958RAJ   Epstein-Barr virus, artifactual joining of B95-8 c

 LOCUS       HS4B958RAJ 184113 bp ds-DNA             VRL       23-DEC-1994
 DEFINITION  Epstein-Barr virus, artifactual joining of B95-8 complete genome
             and the sequences from Raji of the large deletion found in B95-8.
 ACCESSION   M80517
 KEYWORDS    .
 SOURCE      Human herpesvirus 4 DNA.
   ORGANISM  Human herpesvirus 4
             Viridae; ds-DNA enveloped viruses; Herpesviridae;
             Gammaherpesvirinae.
 REFERENCE   1  (sites)
   AUTHORS   Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J.,
             Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C.,
             Tuffnell,P.S. and Barrell,B.G.
   TITLE     DNA sequence and expression of the B95-8 Epstein-Barr virus genome
   JOURNAL   Nature 310, 207-211 (1984)
   STANDARD  full automatic
 REFERENCE   2  (sites)
   AUTHORS   Parker,B.D., Bankier,A., Satchwell,S., Barrell,B. and Farrell,P.J.
   TITLE     Sequence and transcription of Raji Epstein-Barr virus DNA spanning
             the B95-8 deletion region
   JOURNAL   Virology 179, 339-346 (1990)
   STANDARD  full automatic
 REFERENCE   3  (sites)
   AUTHORS   Sample,J., Brooks,L., Sample,C., Young,L., Rowe,M., Gregory,C.,
             Rickinson,A. and Kieff,E.
   TITLE     Restricted Epstein-Barr virus protein expression in Burkitt
             lymphoma is due to a different Epstein-Barr nuclear antigen 1
             transcriptional initiation site
   JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 6343-6347 (1991)
   STANDARD  full automatic
 REFERENCE   4  (bases 1 to 184113)
   AUTHORS   Jenson,H.B.
   TITLE     GenBank Curator Program
   JOURNAL   Unpublished (1992)
   STANDARD  full automatic
 COMMENT     The B95-8 genome (V01555) has a large deletion in the right side of
             the genome which has been sequenced in Raji (M35547).  These
             sequences have been joined to form an extended and more complete,
             although artifactual, EBV sequence.
             For features, refer to feature tables of V01555 and M35547.
```



```
                NCBI gi: 330330
 FEATURES             Location/Qualifiers
      source          1..184113
                      /organism="Human herpesvirus 4"
                      /sequenced_mol="DNA"
      misc_feature    1..152008
                      /note="B95-8 sequences (corresponds to 1-152,008 of
                      V01555)"
      misc_feature    152009..152012
                      /note="Overlap of B95-8 and Raji sequences at B95-8
                      deletion point (corresponds to 152,009-152,012 in V01555,
                      and 1-4 in M35547)"
      misc_feature    153013..163839
                      /note="Raji sequences (corresponds to 5-11,831 of M35547)"
      misc_feature    163840..163843
                      /note="Overlap of B95-8 and Raji sequences at B95-8
                      deletion point (corresponds to 152,009-152,012 of V01555,
                      and 11,832-11,835 of M35547)"
      misc_feature    163844..184113
                      /note="B95-8 sequences (corresponds to 152,013-172,282 of
                      V01555)"
 BASE COUNT    36002 a  55824 c  54622 g  37665 t
 ORIGIN

Initial Score    =    170  Optimized Score  =    454  Significance = 10.52
Residue Identity =    50%  Matches          =    529  Mismatches   =   418
Gaps             =    105  Conservative Substitutions               =     0

             X       10          20        30        40        50
             CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
    CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
107030    107040    107050    107060    107070    107080    107090    107100

    60        70          80        90          100       110       120
    CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
    GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
       107110    107120    107130    107140    107150    107160    107170

      130       140       150       160       170       180       190
    CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
    CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
          107180    107190    107200    107210    107220    107230

      200       210       220       230       240       250       260
    -TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
    GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
107240    107250    107260    107270    107280    107290    107300

          270       280       290       300       310       320
    CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
    GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
107310    107320    107330    107340    107350    107360    107370

     330       340       350       360       370       380       390
```

```
CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
  ||       ||  | |||||  |  |||  || |  ||||| ||||||  |   | ||||||  ||||| ||
GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
 107380    107390     107400     107410    107420    107430    107440

  400       410       420       430       440       450       460
GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
|||| ||| ||||||   | ||||| |  |  ||  | ||    |||| || || |||    || | ||||
GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
  107450    107460    107470    107480    107490    107500    107510

       470         480       490       500       510       520
GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
 | ||  ||  ||||   ||  ||| | |     ||||  ||||    |||  | |  | || | |  |
AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
  107520    107530    107540    107550    107560    107570    107580

530       540       550       560       570       580       590       600
CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
  ||||   || | ||  ||| |   |  || |    |||||||| |||  | |  || | ||| |    |
AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
107590    107600       107610    107620    107630    107640    107650

       610         620           630       640       650       660
AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
|| |||  ||||  || ||  ||  ||||    ||   || | ||  | | |     |    | | ||
AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
  107660    107670    107680    107690      107700    107710    107720

 670       680       690       700       710       720       730
AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
||||  |    |||    || |   |||  ||  | |  |||  ||    | |  | |  | | |||||    |
AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
  107730    107740    107750    107760    107770    107780    107790

     740       750       760       770       780           790
AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA----CTGGGTAATCTT
||  ||  ||||      ||| |||   |  |  | | |||  | |||  | ||  |    | ||      |
AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
107800    107810    107820    107830      107840    107850    107860

800       810       820       830       840       850       860
  GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
  || |  || ||   |||   |  || |||    | || | | ||  ||||| |  || || |    || |||
  GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
      107870    107880      107890    107900    107910    107920

 870       880       890       900       910       920       930
  AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
  |||||     ||   | | | ||  ||||    ||||  |  || || |  ||  |||| | |||   || |
  AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
07930    107940    107950      107960    107970    107980    107990

940       950       960          970       980       X
  GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
  | | |  | | ||  |   || | |||    || |  | ||   || |  |
  GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
108000    108010    108020    108030    108040    108050
```


9. US-08-162-407-5' (1-988)
   HS4B958RAJ Epstein-Barr virus, artifactual joining of B95-8 c

```
LOCUS       HS4B958RAJ 184113 bp ds-DNA             VRL       03-OCT-1991
DEFINITION  Epstein-Barr virus, artifactual joining of B95-8 complete genome
            and the sequences from Raji of the large deletion found in B95-8.
ACCESSION   M80517
KEYWORDS    .
SOURCE      Epstein-Barr virus DNA.
  ORGANISM  Epstein-Barr virus
            Viridae; ds-DNA enveloped viruses; Herpesviridae;
            Gammaherpesvirinae.
REFERENCE   1  (sites)
  AUTHORS   Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J.,
            Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C.,
            Tuffnell,P.S. and Barrell,B.G.
  TITLE     DNA sequence and expression of the B95-8 Epstein-Barr virus genome
  JOURNAL   Nature 310, 207-211 (1984)
  STANDARD  full automatic
REFERENCE   2  (sites)
  AUTHORS   Parker,B.D., Bankier,A.T., Satchwell,S.C., Barrell,B.G. and
            Farrell,P.J.
  TITLE     Sequence and transcription of Raji Epstein-Barr virus DNA spanning
            the B95-8 deletion region
  JOURNAL   Virology 179, 339-346 (1990)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 184113)
  AUTHORS   Jenson,H.B.
  TITLE     GenBank Curator Program
  JOURNAL   Unpublished (1992)
  STANDARD  full automatic
COMMENT     The B95-8 genome (V01555) has a large deletion in the right side of
            the genome which has been sequenced in Raji (M35547).  These
            sequences have been joined to form an extended and more complete,
            although artifactual, EBV sequence.
            For features, refer to feature tables of V01555 and M35547.

            NCBI gi: 330330
FEATURES             Location/Qualifiers
     source          1..184113
                     /organism="Epstein-Barr virus"
                     /sequenced_mol="DNA"
     misc_feature    1..152008
                     /note="B95-8 sequences (corresponds to 1-152,008 of
                     V01555)"
     misc_feature    152009..152012
                     /note="Overlap of B95-8 and Raji sequences at B95-8
                     deletion point (corresponds to 152,009-152,012 in V01555,
                     and 1-4 in M35547)"
     misc_feature    153013..163839
                     /note="Raji sequences (corresponds to 5-11,831 of M35547)"
     misc_feature    163840..163843
                     /note="Overlap of B95-8 and Raji sequences at B95-8
                     deletion point (corresponds to 152,009-152,012 of V01555,
                     and 11,832-11,835 of M35547)"
     misc_feature    163844..184113
                     /note="B95-8 sequences (corresponds to 152,013-172,282 of
                     V01555)"
BASE COUNT    36002 a  55824 c  54622 g  37665 t
ORIGIN
```

```
Initial Score    =    170  Optimized Score  =    454  Significance = 10.52
Residue Identity =    50%  Matches          =    529  Mismatches   =   418
Gaps             =    105  Conservative Substitutions               =     0

            X       10          20          30          40          50
            CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
            ||| |||  ||    ||| ||  | |    ||    |  ||   ||     || |  |    |
    CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
107030    107040    107050    107060    107070    107080    107090    107100

    60        70          80          90          100         110         120
    CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
     |    | |||||  | | ||  ||| ||||  ||  |    ||| | |   | || ||      |  |||||
    GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
        107110    107120    107130    107140    107150    107160    107170

       130       140         150         160         170         180         190
    CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
    | |   | ||  ||      |     | |  | | ||  || |  ||   ||  |  || || | | | | ||
    CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
          107180    107190    107200      107210    107220    107230

      200       210         220         230         240         250         260
    -TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
      ||  |   |     |   | |    ||| |    |||||| || || |  |  ||||   ||  ||
    GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
107240    107250      107260    107270    107280    107290    107300

          270         280         290         300         310         320
    CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
      |   |||| |||  | |  ||| | ||    ||| | ||||||   |     ||| | || | | |    |
    GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
107310    107320    107330    107340    107350          107360    107370

     330       340         350         360         370         380         390
    CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
      ||        ||  | |||||  |  |||  || |  ||||| ||||||  |    | ||||||  |||||  ||
    GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
     107380    107390      107400      107410    107420    107430    107440

       400       410         420         430         440         450         460
    GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
    |||| ||| ||||||   | |||||  |  |  ||  | ||     |||| || || |||    || | ||||
    GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
      107450    107460    107470    107480      107490    107500    107510

          470         480         490         500         510         520
    GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
     | ||  ||  ||||   ||  ||| | |     ||||   ||||    |||  | |  | || | | |  |
    AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
      107520    107530    107540    107550    107560    107570    107580

   530       540         550         560         570         580         590         600
    CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
      ||||   || | ||  ||| |   |  || |    |||||||| ||| | |  || | ||| |    |
    AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
    107590    107600      107610    107620    107630    107640    107650

          610         620         630         640         650         660
```



```
    AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
    || |||  ||||  || ||  ||  ||||    ||   || | ||  | | |     |     | | ||
    AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
      107660    107670    107680    107690      107700    107710    107720

     670       680         690         700         710         720         730
    AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
    |||| |    |||     || |   |||  ||  | | |||  ||   | |  | | | | |||||    |
    AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
      107730    107740    107750    107760    107770    107780    107790

         740       750         760         770         780         790
    AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA----CTGGGTAATCTT
    ||  ||  ||||     ||| |||   |  |  | | |||  | |||  | ||  |    | ||     |
    AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
    107800    107810    107820    107830      107840    107850    107860

  800       810         820         830         840         850         860
    GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
    || |  || ||   |||   |  || |||    | || | | ||  |||||  |  || || |    || |||
    GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
        107870    107880      107890    107900    107910    107920

   870       880         890         900         910         920         930
    AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
    |||||     ||   | | | ||  ||||    ||||  |  || || |  ||  |||| | |||   || |
    AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
07930    107940    107950      107960    107970    107980    107990

  940       950         960         970         980         X
    GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCCGGAA-TTCCGGCCG
    | | | |   | | ||  |   || | |||    || |  | ||   || |  |
    GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
  108000    108010    108020    108030    108040    108050
```

10. US-08-162-407-5' (1-988)
    EBV Epstein-Barr virus (EBV) genome. The complete sequ

```
LOCUS       EBV        172281 bp    DNA   circular  VRL       12-SEP-1993
DEFINITION  Epstein-Barr virus (EBV) genome. The complete sequence [1-10] was
            determined from DNA from B95-8 cells cloned by Arrand et al [11].
            B95-8 is a productive marmoset lymphoblastoid cell line
            immortalized with human EBV from a mononucleosis patient.
ACCESSION   V01555 J02070 K01729 K01730 V01554 X00498 X00499 X00784
KEYWORDS    DNA polymerase; EBNA; genome; ribonucleotide reductase;
            tandem repeat; terminal repeat.
SOURCE      Human herpesvirus 4.
  ORGANISM  Human herpesvirus 4
            Viridae; ds-DNA enveloped viruses; Herpesviridae;
            Gammaherpesviridae.
REFERENCE   1  (bases 1 to 172281)
  AUTHORS   Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J.,
            Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C.,
            Tuffnell,P.S. and Barrell,B.G.
  TITLE     DNA sequence and expression of the B95-8 Epstein-Barr virus genome
  JOURNAL   Nature 310, 207-211 (1984)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 172281)
  AUTHORS   Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.
```

```
  TITLE     Sequence analysis and in vitro transcription of portions of the
            Epstein-Barr virus genome
  JOURNAL   J. Cell. Biochem. 19, 267-274 (1982)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 172281)
  AUTHORS   Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.
  TITLE     Homologous upstream sequences near Epstein-Barr virus promoters
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 80, 1565-1569 (1983)
  STANDARD  full automatic
REFERENCE   4  (bases 1 to 172281)
  AUTHORS   Farrell,P.J., Bankier,A.T., Seguin,C., Deininger,P.L. and
            Barrell,B.G.
  TITLE     Latent and lytic cycle promoters of Epstein-Barr virus
  JOURNAL   EMBO J. 2, 1331-1338 (1983)
  STANDARD  full automatic
REFERENCE   5  (bases 142687 to 159853)
  AUTHORS   Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.
  TITLE     Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8
            Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1, 21-45 (1983)
  STANDARD  full automatic
REFERENCE   6  (bases 112620 to 125316)
  AUTHORS   Seguin,C., Farrell,P.J. and Barrell,B.G.
  TITLE     DNA sequence and transcription of the BamHI fragment B region of
            B95-8 Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1, 369-392 (1983)
  STANDARD  full automatic
REFERENCE   7  (bases 159853 to 172281)
  AUTHORS   Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J.
            and Barrell,B.G.
  TITLE     DNA sequence analysis of the EcoRI Dhet fragment of B95-8
            Epstein-Barr virus containing the terminal repeat sequences
  JOURNAL   Mol. Biol. Med. 1, 425-445 (1983)
  STANDARD  full automatic
REFERENCE   8  (bases 87650 to 92703)
  AUTHORS   Biggin,M., Farrell,P.J. and Barrell,B.G.
  TITLE     Transcription and DNA sequence of the BamHI L fragment of B95-8
            Epstein-Barr virus
  JOURNAL   EMBO J. 3, 1083-1090 (1984)
  STANDARD  full automatic
REFERENCE   9  (bases 76089 to 79808)
  AUTHORS   Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.
  TITLE     Homology between two EBV early genes and HSV ribonucleotide
            reductase and 38K genes
  JOURNAL   Nucleic Acids Res. 12, 5087-5099 (1984)
  STANDARD  full automatic
REFERENCE   10 (bases 1 to 172281)
  AUTHORS   Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeoch,D.
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   11 (bases 1 to 172281)
  AUTHORS   Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and
            Griffin,B.E.
  TITLE     Molecular cloning of the complete Epstein-Barr virus genome as a
            set of overlapping restriction endonuclease fragments
  JOURNAL   Nucleic Acids Res. 9, 2999-3014 (1981)
  STANDARD  full automatic
REFERENCE   12 (bases 1 to 172281)
  AUTHORS   Kozak,M.
  TITLE     Possible role of flanking nucleotides in recognition of the AUG
```



```
            initiator codon by eukaryotic ribosomes
  JOURNAL   Nucleic Acids Res. 9, 5233-5262 (1981)
  STANDARD  full automatic
REFERENCE   13 (bases 7315 to 9312)
  AUTHORS   Yates,J., Warren,N., Reisman,D. and Sugden,B.
  TITLE     A cis-acting element from the Epstein-Barr viral genome that
            permits stable replication of recombinant plasmids in latently
            infected cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 81, 3806-3810 (1984)
  STANDARD  full automatic
REFERENCE   14 (bases 45415 to 52824)
  AUTHORS   Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.
  TITLE     The EB virus genome in Daudi Burkitt's lymphoma cells has a
            deletion similar to that observed in a non-transforming strain
            (P3HR-1) of the virus
  JOURNAL   EMBO J. 3, 813-821 (1984)
  STANDARD  full automatic
REFERENCE   15 (bases 45644 to 52450)
  AUTHORS   Jeang,K.T. and Hayward,S.D.
  TITLE     Organization of the Epstein-Barr virus DNA molecule. III. Location
            of the P3HR-1 deletion junction and characterization of the NotI
            repeat units that form part of the template for an abundant
            12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
  JOURNAL   J. Virol. 48, 135-148 (1983)
  STANDARD  full automatic
REFERENCE   16 (bases 1 to 172281)
  AUTHORS   Farrell,P.J. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-1984) to the EMBL/GenBank/DDBJ databases
  STANDARD  full automatic
REFERENCE   17 (bases 1 to 172281)
  AUTHORS   Bodescot,M. and Perricaudet,M.
  TITLE     Clustered alternative splice sites in Epstein-Barr virus RNAs
  JOURNAL   Nucleic Acids Res. 15, 5887 (1987)
  STANDARD  full automatic
REFERENCE   18 (bases 1 to 172281)
  AUTHORS   Laux,G., Perricaudet,M. and Farrell,P.J.
  TITLE     A spliced Epstein-Barr virus gene expressed in immortalized
            lymphocytes is created by circularization of the linear viral
            genome
  JOURNAL   EMBO J. 7, 769-774 (1988)
  STANDARD  full automatic
REFERENCE   19 (bases 1 to 172281)
  AUTHORS   Farrell,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1988) to the EMBL/GenBank/DDBJ databases. Farrell
            P., Ludwig Institute for Cancer Research, St. Mary's Hospital
            Medical School, Norfolk Place London W2 1PG
  STANDARD  full automatic
COMMENT     CDS

            Listed under this feature are all known protein coding regions as
            well as all the major open reading frames in the sequence. In
            general the term major is taken as the longest frame in a
            particular region taking into account the adjacent longest frames
            and likely transcription signals. Note that on this basis some long
            overlapping frames have been excluded and on the other hand some
            small frames have been included which might represent exons or
            genes because they occur in a logical combination with other
            features or because of some other experimental data. The reading
```

frames are named according to the Bam H1 fragment in which they start. eg BALF3 is the third leftward frame starting in Bam H1 fragment A. BORF1 is the first rightward frame in Bam H1 fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITEs of POLYA signals

This feature lists all occurences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog ATTAAA is only listed when it is found in a position close to the end of a major reading frame.

SITEs of DONOR and ACCEPT sequences

This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case.

Restriction enzyme SITEs.

Only the positions of the sites Bam HI (BAM) are listed.

RPT

This feature is used to define repetitive sequences.

SITE DEL

This feature defines deletions in B95-8 with respect to other strains such as RAJI and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN

Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurences noted.

ORGRPL

Denotes the region that encompasses an origin of replication (ori P).[13].

NUMBERING

The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has benn moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

NCBI gi: 59074

```
FEATURES             Location/Qualifiers
     source          1..172281
                     /organism="Epstein-Barr virus"
                     /strain="B95-8"
     mRNA            58..272
                     /note="exon 2 terminal protein RNA"
     mRNA            360..458
                     /note="exon 3 terminal protein RNA"
     misc_feature    complement(535)
                     /note="polyA signal: AATAAA"
     mRNA            540..788
                     /note="exon 4 terminal protein RNA"
     mRNA            871..951
                     /note="exon 5 terminal protein RNA"
     mRNA            1026..1196
                     /note="exon 6 terminal protein RNA"
     promoter        complement(1192)
                     /note="TATA: TATAAAT"
     mRNA            1280..1495
                     /note="exon 7 terminal protein RNA"
     promoter        complement(1383)
                     /note="TATA: CATAAAA"
     mRNA            1574..1682
                     /note="exon 8 terminal protein RNA"
     promoter        1676
                     /note="TATA: TATAAAG"
     promoter        1691
                     /note="TATA: TATTAAA BN-R1 late promoter before BNRF1,
                     gives 4.1kb late RNA. Probably encodes non glycosylated
                     140kd protein in membrane antigen. Also two latent RNAs
                     spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson
                     et al, 1985). The longer one encodes terminal protein."
     CDS             1736..5692
                     /note="BNRF1 reading frame, 5 NXT/S;  NCBI gi: 59075"
                     /codon_start=1
                     /translation="MEERGRETQMPVARYGGPFIMVRLFGQDGEANIQEERLYELLSD
                     PRSALGLDPGPLIAENLLLVALRGTNNDPRPQRQERARELALVGILLGNGEQGEHLGT
                     ESALEASGNNYVYAYGPDWMARPSTWSAEIQQFLRLLGATYVLRVEMGRQFGFEVHRS
                     RPSFRQFQAINHLVLFDNALRKYDSGQVAAGFQRALLVAGPETADTRPDLRKLNEWVF
                     GGRAAGGRQLADELKIVSALRDTYSGHLVLQPTETLDTWKVLSRDTRTAHSLEHGFIH
                     AAGTIQANCPQLFMRRQHPGLFPFVNAIASSLGWYYQTATGPGADARAAARRQQAFQT
                     RAAAECHAKSGVPVVAGFYRTINATLKGGEGLQPTMFNGELGAIKHQALDTVRYDYGH
                     YLIMLGPFQPWSGLTAPPCPYAESSWAQAAVQTALELFSALYPAPCISGYARPPGPSA
```

```
                     VIEHLGSLVPKGGLLLFLSHLPDDVKDGLGEMGPARATGPGMQQFVSSYFLNPACSNV
                     FITVRQRGEKINGRTVLQALGRACDMAGCQHYVLGSTVPLGGLNFVNDLASPVSTAEM
                     MDDFSPFFTVEFPPIQEEGASSPVPLDVDESMDISPSYELPWLSLESCLTSILSHPTV
                     GSKEHLVRHTDRVSGGRVAQQPGVGPLDLPLADYAFVAHSQVWTRPGGAPPLPYRTWD
                     RMTEKLLVSAKPGGENVKVSGTVITLGEQGYKVSLDLREGTRLAMAEALLNAACAPIL
                     DPEDVLLTLHLHLDPRRADNSAVMEAMTAASDYARGLGVKLTFGSASCPETGSSASNF
                     MTVVASVSAPGEFSGPLITPVLQKTGSLLIAVRCGDGKIQGGSLFEQLFSDVATTPRA
                     PEALSLKNLFRAVQQLVKSGIVLSGHDISDGGLVTCLVEMALAGQRGVTITMPVASDY
                     LPEMFAEHPGLVFEVEERSVGEVLQTLRSMNMYPAVLGRVGEQGPDQMFEVQHGPETV
                     LRQSLRLLLGTWSSFASEQYECLRPDRINRSMHVSDYGYNEALAVSPLTGKNLSPRRL
                     VTEPDDPRCQVAVLCAPGTRGHESLLAAFTNAGCLCRRVFFREVRDNTFLDKYVGLAIG
                     GVHGARDSALAGRATVALINRFPALRDAILKFLNRPDTFSVALGELGVQVLAGLGAVG
                     STDNPPAPGVEVNVQRSPLILAPNASGMFESRWLNISIPATTSSVMLRGLRGCVLPCW
                     VQGSCLGLQFTNLGMPYVLQNAHQIACHFHSNGTDAWRFAMNYPRNPTEQGNIAGLCS
                     RDGRHLALLCDPSLCTDFWQWEHIPPAFGHPTGCSPWTLMFQAAHLWSLRHGRPSE"
misc_feature         complement(1795)
                     /note="polyA signal: AATAAA"
misc_feature         3955
                     /note="BAM: Bam H1 Nhet/h"
misc_feature         3994
                     /note="BAM: Bam H1 h/C"
mRNA                 5408..5856
                     /note="exon 9 terminal protein RNA"
misc_feature         5841
                     /note="polyA signal: AATAAA, end of 4.1kb late RNA and TP
                     latent RNA."
misc_feature         5863
                     /note="alternative end to TP cDNAs"
promoter             6097
                     /note="TATA: TATAAGA"
misc_feature         6629..6795
                     /note="Pol III RNA EBER 1"
promoter             complement(6823)
                     /note="TATA: CATAAAT"
misc_feature         6956..7128
                     /note="Pol III RNA EBER 2"
rep_origin           7315..9312
                     /note="origin of replication, ori P (Yates et al, 1984,
                     1985)"
repeat_region        7421..8042
                     /note="21x30bp repeats, binding sites for EBNA-1 (site I,
                     Rawlins et al, 1985). Tandem repeat part of oriP (Reisman
                     et al, 1985). Also functions as a cell type specific
                     enhancer (Reisman et al, 1985; Lupton and Levine, 1985)"
promoter             7738
                     /note="TATA: TATAAAT"
promoter             7888
                     /note="TATA: TATAAAT"
promoter             8573
                     /note="TATA: CATAAAT"
misc_feature         complement(8680)
                     /note="polyA signal: AATAAA"
misc_feature         complement(8755)
                     /note="polyA signal: AATAAA"
misc_feature         8962
                     /note="polyA signal: AATAAA"
misc_feature         9021..9133
                     /note="HPN: dyad symmetry, site II for EBNA-1 binding
                     (Rawlins et al, 1985). Dyad symmetry part of oriP (Reisman
                     et al, 1985)"
```



```
promoter             complement(9398)
                     /note="TATA: TATAAAT"
promoter             9631
                     /note="TATA: TATAAAT BC-R1 late promoter before BCRF1"
CDS                  9675..10187
                     /note="BCRF1 reading frame;  NCBI gi: 59076"
                     /codon_start=1
                     /translation="MERRLVVTLQCLVLLYLAPECGGTDQCDNFPQMLRDLRDAFSRV
                     KTFFQTKDEVDNLLLKESLLEDFKGYLGCQALSEMIQFYLEEVMPQAENQDPEAKDHV
                     NSLGENLKTLRLRLRRCHRFLPCENKSKAVEQIKNAFNKLQEKGIYKAMSEFDIFINY
                     IEAYMTIKAR"
promoter             10076
                     /note="TATA: GATAAAA"
misc_feature         complement(10148)
                     /note="polyA signal: AATAAA"
misc_feature         10173
                     /note="polyA signal: ATTAAA"
misc_feature         10257
                     /note="polyA signal: AATAAA, end of 0.8kb late RNA from
                     BCR1 and end of 1.6 kb late RNA, start unknown"
misc_feature         complement(10277)
                     /note="polyA signal: AATAAA"
promoter             complement(10975)
                     /note="TATA: CATAAAT"
promoter             11305
                     /note="TATA: TACAAAA; BCR2 promoter for highly spliced
                     EBNA latent RNAs."
mRNA                 11336..11480
                     /note="exon C1 of Bodescot et al (1986) RNAs"
promoter             11524
                     /note="TATA: TATAATT"
misc_feature         complement(11587)
                     /note="polyA signal: AATAAA"
promoter             complement(11606)
                     /note="TATA: CATAAAT"
mRNA                 11626..11657
                     /note="exon C2 of Bodescot et al (1986) RNAs"
promoter             11796
                     /note="TATA: TATAAGT"
promoter             complement(11799)
                     /note="TATA: TATAAAA"
repeat_region        12001..15072
                     /note="3072 repeat 1"
CDS                  12541..13692
                     /note="BCRF2 3072 repeat, reading frame 1;  NCBI gi:
                     583873"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         13215
                     /note="BAM: BamH1 C/W"
promoter             14352
                     /note="*TATA: TATAAAG BWR1 one of the promoters for highly
                     spliced EBNA and LP RNAs (Sample et al, 1986; Speck et al,
                     1986)"
```

```
mRNA            14384..14410
                /note="*exon W0 of EBNA/LP RNAs"
mRNA            14554..14619
                /note="*exon W1 (also W66) part of leader protein (LP)
                gene. LP is also called EBNA-5 (Dillner et al, 1986) and
                EBNA-4 (Rowe et al, 1987)."
mRNA            14559..14619
                /note="*exon W1' (also W61) of EBNA/LP RNAs forms
                initiator met when fused to exon W0 or exon C2."
mRNA            14701..14832
                /note="*exon W2 (also W132) part of LP gene"
repeat_region   15073..18144
                /note="3072 repeat 2"
CDS             15613..16764
                /note="BWRF1 reading frame 2;  NCBI gi: 583874"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    16287
                /note="BAM: BamH1 W/W"
promoter        17424
                /note="TATA: TATAAAG"
mRNA            17626..17691
                /note="Exon W1"
mRNA            17773..17904
                /note="Exon W2"
repeat_region   18145..21216
                /note="3072 repeat 3"
CDS             18685..19836
                /note="BWRF1 reading frame 3;  NCBI gi: 583875"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    19359
                /note="BAM: BamH1 W/W"
promoter        20496
                /note="TATA: TATAAAG"
mRNA            20698..20763
                /note="Exon W1"
mRNA            20845..20976
                /note="Exon W2"
repeat_region   21217..24288
                /note="3072 repeat 4"
CDS             21757..22908
                /note="BWRF1 reading frame 4;  NCBI gi: 583876"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
```



```
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    22431
                /note="BAM: BamH1 W/W"
promoter        23568
                /note="TATA: TATAAAG"
mRNA            23771..23835
                /note="Exon W1"
mRNA            23917..24048
                /note="Exon W2"
repeat_region   24289..27360
                /note="3072 repeat 5"
CDS             24829..25980
                /note="BWRF1 reading frame 5;  NCBI gi: 583877"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    25503
                /note="BAM: BamH1 W/W"
promoter        26640
                /note="TATA: TATAAAG"
mRNA            26842..26907
                /note="Exon W1"
mRNA            26989..27120
                /note="Exon W2"
repeat_region   27361..30432
                /note="3072 repeat 6"
CDS             27901..29052
                /note="BWRF1 reading frame 6;  NCBI gi: 583878"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    28575
                /note="BAM: BamH1 W/W"
promoter        29712
                /note="TATA: TATAAAG"
mRNA            29914..29979
                /note="Exon W1"
mRNA            30061..30192
                /note="Exon W2"
repeat_region   30433..33504
                /note="3072 repeat 7"
CDS             30973..32124
                /note="BWRF1 reading frame 7;  NCBI gi: 583879"
                /codon_start=1
                /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
```

```
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         31647
                     /note="BAM: BamH1 W/W"
promoter             32784
                     /note="TATA: TATAAAG"
mRNA                 32986..33051
                     /note="Exon W1"
mRNA                 33133..33264
                     /note="Exon W2"
repeat_region        33505..36576
                     /note="3072 repeat 8"
CDS                  34045..35196
                     /note="BWRF1 reading frame 8;  NCBI gi: 583880"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         34719
                     /note="BAM: BamH1 W/W"
promoter             35856
                     /note="TATA: TATAAAG"
mRNA                 36058..36123
                     /note="Exon W1"
mRNA                 36205..36336
                     /note="Exon W2"
repeat_region        36577..39648
                     /note="3072 repeat 9"
CDS                  37117..38268
                     /note="BWRF1 reading frame 9;  NCBI gi: 583881"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         37791
                     /note="BAM: BamH1 W/W"
promoter             38928
                     /note="TATA: TATAAAG"
mRNA                 39130..39195
                     /note="Exon W1"
mRNA                 39277..39408
                     /note="Exon W2"
repeat_region        39649..42720
                     /note="3072 repeat 10"
CDS                  40189..41340
                     /note="BWRF1 reading frame 10;  NCBI gi: 583882"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
```



```
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         40863
                     /note="BAM: BamH1 W/W"
promoter             42000
                     /note="TATA: TATAAAG"
mRNA                 42202..42267
                     /note="Exon W1"
mRNA                 42349..42480
                     /note="Exon W2"
repeat_region        42721..45792
                     /note="3072 repeat 11"
CDS                  43261..44412
                     /note="BWRF1 reading frame 11;  NCBI gi: 583883"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         43935
                     /note="BAM: BamH1 W/W"
promoter             45072
                     /note="TATA: TATAAAG"
mRNA                 45274..45339
                     /note="Exon W1"
misc_feature         45415..52824
                     /note="DEL: DAUDI deletion (Jones et al, 1984)"
mRNA                 45421..45552
                     /note="Exon W2"
misc_feature         45644..52450
                     /note="DEL: P3HR1 deletion (Jeang and Hayward, 1983)"
repeat_region        45793..47643
                     /note="3072 repeat 12"
CDS                  46333..47484
                     /note="BWRF1 reading frame 12;  NCBI gi: 583884"
                     /codon_start=1
                     /translation="XWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                     PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                     TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                     PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                     HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                     LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                     VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature         47007
                     /note="BAM: BamH1 W/Y"
mRNA                 47761..47793
                     /note="Exon Y1 Bodescot et al, 1984"
promoter             47831
                     /note="TATA: TATAAGT"
mRNA                 47878..47999
                     /note="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and
                     Strominger,1985), last common exon"
misc_feature         complement(48023)
                     /note="polyA signal: AATAAA"
CDS                  48386..50032
                     /note="Coding exon for EBNA-2 (Sample et al,1986)"
```

```
                     /codon_start=1
mRNA                 48386..48444
                     /note="exon Bodescot et al, 1984"
CDS                  48429..49967
                     /note="BYRF1, encodes EBNA-2 (Dambaugh et al, 1984;
                     Dillner et al, 1984);  NCBI gi: 583885"
                     /codon_start=1
                     /translation="XCGWAGTCQQPSKHPRLCFALSAAIMPTFYLALHGGQTYHLIVD
                     TDSLGNPSLSVIPSNPYQEQLSDTPLIPLTIFVGENTGVPPPLPPPPPPPPPPPPPPP
                     PPPPPPPPPPPSPPPPPPPPPPPQRRDAWTQEPSPLDRDPLGYDVGHGPLASAMRMLW
                     MANYIVRQSRGDRGLILPQGPQTAPQARLVQPHVPPLRPTAPTILSPLSQPRLTPPQP
                     LMMPPRPTPPTPLPPATLTVPPRPTRPTTLPPTPLLTVLQRPTELQPTPSPPRMHLPV
                     LHVPDQSMHPLTHQSTPNDPDSPEPRSPTVFYNIPPMPLPPSQLPPPAAPAQPPPGVI
                     NDQQLHHLPSGPPWWPPICDPPQPSKTQGQSRGQSRGRGRGRGRGRGKGKSRDKQRKP
                     GGPWRPEPNTSSPSMPELSPVLGLHQGQGAGDSPTPGPSNAAPVCRNSHTATPNVSPI
                     HEPESHNSPEAPILFPDDWYPPSIDPADLDESWDYIFETTESPSSDEDYVEGPSKRPR
                     PSIQ"
repeat_region        48678..48800
                     /note="14 x CCCCCACCA repeats"
misc_feature         48848
                     /note="BAM: BamH1 Y/H"
promoter             49350
                     /note="TATA: TATAACA"
promoter             complement(49353)
                     /note="TATA: TATAAAA"
repeat_region        49525..49578
                     /note="9 x GGGGCA repeats"
mRNA                 49852..50032
                     /note="exon (Bodescot et al 1984)"
misc_feature         50003
                     /note="polyA signal: AATAAA, end of Bodescot T1 RNA and
                     EBNA-2 RNA (3.0kb latent RNA in IB4 cells)"
promoter             complement(50156)
                     /note="TATA: TATAAGT"
misc_feature         complement(50317)
                     /note="polyA signal: AATAAA, end of 2.5kb early RNA from
                     52817"
repeat_region        50578..52115
                     /note="12 x 125bp repeats"
misc_feature         complement(50578..52557)
                     /note="BHLF1 early reading frame"
misc_feature         52654..53697
                     /note="region homologous to Eco R1 C of Raji"
promoter             complement(52817)
                     /note="TATA: GATAAAA promoter for 2.5kb early RNA
                     containing BHLF1 (Jeang and Hayward, 1983; Freese et
                     al,1983)"
promoter             53759
                     /note="TATA: TATTAAC likely promoter for class III and IV
                     early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature         53895
                     /note="DONOR: CGGGTAACT donor for splice to 54335 in class
                     IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature         54335
                     /note="ACCEPT: TTTTCTAG acceptor from 48444 in class I,
                     47999 in class II, and 53895 in class IV early RNAs
                     encoding BHRF1 (Pearson et al, 1987)"
misc_feature         54376..54948
                     /note="BHRF1 reading frame, limited homolgy to bcl-2 gene.
                     Early gene in B95-8 cells and part of restricted EA
```



```
                     complex."
promoter             54591
                     /note="TATA: TATAACA"
promoter             complement(54594)
                     /note="TATA: TATAAAT"
misc_feature         54853
                     /note="BAM: BamH1 H/F"
misc_feature         complement(54929)
                     /note="polyA signal: AATAAA"
promoter             complement(54977)
                     /note="TATA: TATAAAG"
misc_feature         55518
                     /note="polyA signal: AATAAA, 3' end of 2.5kb, 1.9kb, 1.7kb
                     and 0.6kb early RNAs"
misc_feature         complement(55982..56935)
                     /note="BFLF2 reading frame, 4 NXT/S, homologous to RF 27
                     in VZV and HFRF2 in CMV"
misc_feature         complement(55990)
                     /note="polyA signal: AATAAA, 3' end of 2.3kb and 1.1kb
                     early RNAs from 58568 and 57081"
promoter             complement(56132)
                     /note="TATA: TATAAAG"
CDS                  complement(56948..58525)
                     /note="BFLF1 reading frame, 2 NXT/S homologous to RF 26 in
                     VZV and HFRF1 in CMV;  NCBI gi: 59090"
                     /codon_start=1
                     /translation="MAHKVTSANEPNPLTGKRLSSCPLTRSGVTEVAQIAGRTPKMED
                     FVPWTVDNLKSQFEAVGLLMAHSYLPANAEEGIAYPPLVHTYESLSPASTCRVCDLLD
                     TLVNHSDAPVAFFEDYALLCYYCLNAPRAWISSLITGMDFLHILIKYFPMAGGLDSLF
                     MPSRILAIDIQLHFYICRCFLPVSSSDMIRNANLGYYKLEFLKSILTGQSPANFCFKS
                     MWPRTTPTFLTLPGPRTCKDSQDVPGDVGRGLYTALCCHLPTRNRVQHPFLRAEKGGL
                     SPEITTKADYCGLLLGTWQGTDLLGGPGHHAIGLNAEYSGDELAELALAITRPEAGDH
                     SQGPCLLAPMFGLRHKNASRTICPLCESLGAHPDAKDTLDRFKSLILDSFGNNIKILD
                     RIVFLIKTQNTLLDVPCPRLRAWLQMCTPQDFHKHLFCDPLCAINHSITNPSVLFGQI
                     YPPSFQAFKAALAAGQNLEQGVCDSLITLVYIFKSTQVARVGKTILVDVTKELDVVLR
                     IHGLDLVQSYQTSQVYV"
promoter             complement(57081)
                     /note="TATA: TATTTAA before BFLF2; BFL2 promoter gives
                     1.1kb early RNA"
promoter             complement(58088)
                     /note="TATA: GATAAAA"
promoter             complement(58568)
                     /note="TATA: TATTAAA before BFLF1, BFL1 promoter gives
                     2.3kb early RNA"
promoter             58832
                     /note="TATA: TATAAAA before BFRF1"
CDS                  58891..59901
                     /note="BFRF1 early reading frame, 1 NXT/S, homologous to
                     HFLF4 in CMV;  NCBI gi: 59091"
                     /codon_start=1
                     /translation="MASPEERLLDELNNVIVSFLCDSGSLEVERCSGAHVFSRGSSQP
                     LCTVKLRHGQIYHLEFVYKFLAFKLKNCNYPSSPVFVISNNGLATTLRCFLHEPSGLR
                     SGQSGPCLGLSTDVDLPKNSIIMLGQDDFIKFKSPLVFPAELDLLKSMVVCRAYITEH
                     RTTMQFLVFQAANAQKASRVMDMISDMSQQLSRSGQVEDTGARVTGGGGPRPGVTHSG
                     CLGDSHVRGRGGWDLDNFSEAETEDEASYAPWRDKDSWSESEAAPWKKELVRHPIRRH
                     RTRETRRMRGSHSRVEHVPPETRETVVGGAWRYSWRATPYLARVLAVTAVALLLMFLR
                     WT"
CDS                  59610..61583
                     /note="BFRF2 early reading frame, homologous to HFLF5 in
                     CMV;  NCBI gi: 583886"
```

```
                     /codon_start=1
                     /translation="XRRRGELRSLEGQRLVVGIRGGAVEEGTREAPHPQAPDTRDSPY
                     AREPFTGGTRAPRDPGDGGGGSMALFLARHTLSGTGAGCHGRGPAPDVSEVDLTLQAL
                     GERGFSRLLDLGLACLDLSYVEMREFVVWGRPPASEAAVASTPGSLFRSHSSAYWLSE
                     VERPGGLVRWARSQTSPSSLTLAPHLGPSLLSLSVVTGGGCGAVAFCNAFFLAYFLVV
                     RSVFPAFSDRIAAWICDRSPFCENTRAVARGYRGLVKRFLAFVFERSSYDPPLLRQNS
                     RPVERCFAIKNYVPGLDSQSCVTVPSFSRWAQSHASELDPREIRDRVTPATAPSFVAD
                     HASALLASLQKKASDTPCGNPIQWMWYRLLVNSCLRSAHCLLPIPAVSEGGRKTGGGV
                     GEELVGAGGPCLSRDVFVAIVSRNVLSCLLNVPAAGPRAYKCFRSHASRPVSGPDYPP
                     LAVFCMDCGYCLNFGKQTGVGGRLNSFRPTLQFYPRDQKEKHVLTCHASGRVYCSNCG
                     SAAVGCQRLAEPPSARSGWRPRIRAVLPHNAAYELDRGSRLLDAIIPCLGPDRTCMRP
                     VVLRGVTVRQLLYLTLRTEARAVCSICQQRQAPEDARDEPHLFSSCLEVELPPGERCA
                     GCRLYQTRYGTPAAQAHPPGEAGGGFSRQSPAS"
promoter             complement(61062)
                     /note="TATA: GATAAAA"
promoter             61344
                     /note="TATA: TATTTAA before BFRF3"
CDS                  61456..62037
                     /note="BFRF3 early reading frame;  NCBI gi: 583887"
                     /codon_start=1
                     /translation="XCHLVSGVRAAVSIRRVMARRLPKPTLQGRLEADFPDSPLLPKF
                     QELNQNNLPNDVFREAQRSYLVFLTSQFCYEEYVQRTFGVPRRQRAIDKRQRASVAGA
                     GAHAHLGGSSATPVQQAQAAASAGTGALASSAPSTAVAQSATPSVSSSISSLRAATSG
                     ATAAASAAAAVDTGSGGGGQPHDTAPRGARKKQ"
misc_feature         complement(62068)
                     /note="polyA signal: AATAAA"
misc_feature         62069
                     /note="polyA signal: AATAAA, 3' end of 10, 6.5, 3.7, 3.4,
                     3.1, 2.5 and 0.8kb early RNAs"
CDS                  complement(62078..71527)
                     /note="BPLF1 reading frame, 1 NXT/S, analogous to VZV
                     RF22;  NCBI gi: 59094"
                     /codon_start=1
                     /translation="MSNGDWGQSQRTRGTGPVRGIRTMDVNAPGGGSGGSALRILGTA
                     SCNQAHCKFGRFAGIQCVSNCVLYLVKSFLAGRPLTSRPELDEVLDEGARLDALMRQS
                     GILKGHEMAQLTDVPSSVVLRGGGRVHIYRSAEIFGLVLFPAQIANSAVVQSLAEVLH
                     GSYNGVAQFILYICDIYAGAIIIETDGSFYLFDPHCQKDAAPGTPAHVRVSTYAHDIL
                     QYVGAPGAQYTCVHLYFLPEAFETEDPRIFMLEHYGVYDFYEANGSGFDLVGPELVSS
                     DGEAAGTPGADSSPPVMLPFERRIIPYNLRPLPSRSFTSDSFPAARYSPAKTNSPPSS
                     PASAAPASAAPASAAPASAAPASAAPASAAPASAAPASAAPASSPPLFIPIPGLGHTP
                     GVPAPSTPPRASSGAAPQTPKRKKGLGKDSPHKKPTSGRRLPLSSTTDTEDDQLPRTH
                     VPPHRPPSAARLPPPVIPIPHQSPPASPTPHPAPVSTIAPSVTPSPRLPLQIPIPLPQ
                     AAPSNPKIPLTTPSPSPTAAAAPTTTTLSPPPTQQQPPQSAAPAPSPLLPQQQPTPSA
                     APAPSPLLPQQQPPPSAARAPSPLPPQQQPLPSATPAPPPAQQLPPSATTLEPEKNHP
                     PAADRAGTEISPSPPFGQQPSFGDDASGGSGLVRYLSDLEEPFLSMSDSEEAESDLAS
                     DIPTTEDEDMFEDEVFSNSLESGSSAPTSPITLDTARSQYYQTTFDIETPEMDFVPLE
                     SNIARIAGHTYQEQAIVYDPASNREVPEADALSMIDYLLVTVVLEQGLIRSRDRSSVL
                     NLLEFLKDWSGHLQVPTLDLEQLLTSELNIQNLANMLSENKGRAGEFHKHLAAKLEAC
                     LPSLATKDAVRVDAGAKMLAEIPQLAESDDGKFDLEAARRRLTDLLSGGDQEAGEGGG
                     EPEDNSIYRGPHVDVPLVLDDESWKRLLSLAEAARTAVARQQAGVDEEDVRFLALLTA
                     IEYGAPPAASVPPFVHNVAVRSKNAALHVRRCTADIRDKVASAASDYLSYLEDPSLPT
                     VMDFDDLLTHLRHTCQIIASLPLLNIRYTSIEWDYRELLYLGTALSDMSGIPWPLERV
                     EEDDPSIAPLPEFETVAKKQKELETTRENEKRLRTILDDIEAMLGLAGVASAPGAPIS
                     PASPSATPANHDNPEATPPLADTAALTIPVIEKYIANAGSIVGAAKNPTYIRLRDTIQ
                     QIVRSKKYLMNILKSITFYTIDNYIASFEESIDHLYRDLPVLDPEVQDGIDRILDPMV
                     SEALHTFEMGNRLTLEPARLVALQNFATHSTLKETAAAVNLLPGLLAVYDATITGQAP
                     EDALRLLSGLQNQLSQTLIPGKLKKRFLSYLQKLKNNNNDQLRQKEVQAWRLEAEGFK
                     PATEEQLEAFLDTAPNKELKRQYEKKLRQLMETGRKEKEKLREQEDKERQERRAREAN
                     EAWARIRKALGARPEPAPTSPDDWNTLLASLLPDNTDSAAAAAAAVARNTDILDSLTQ
                     ILAAMLLGITRVRRERLRSLLVDDGGAAERMEAAEPGWFTDIETGPLARLDAWPATPA
```



```
                     ATAKEGGGGRGAEEAAGALFRARTAADAIRSALAQTRQALQSPDMKSAVVNTDLEAPY
                     AEYERGLAGLLEKRRAAEAALTAIVSEYVDRTLPEATNDPGQANLPPPPTIPQATAPP
                     RLASDSALWPKKPQLLTRRERDDLLQATGDFFSELLTEAEAAEVRALEEQVRESQTLM
                     AKAHEMAASTRRGFHTALEAVLSRSRDEAPDDELRSLLPSPPKAPVQAPLEAALARAA
                     AGNGSWPYRKSLAAAKWIRGICEAVRGLSEGALALAGGAGAWLNLAAAADGEIHELTR
                     LLEVEGMAQNSMDGMEELRLALATLDPKRVAGGKETVADWKRRLSRLEAIIQEAQEES
                     QLQGTLQDLVTQARGHTDPRQLKIVVEAARGLALGASAGSQYALLKDKLLRYASAKQS
                     FLAFYETAQPTVFVKHPLTNNLPLLITISAPPTGWGNGAPTRRAQFLAAAGPAKYAGT
                     LWLETESPCDPLNPAYVSADTQEPLNYIPVYHNFLEYVMPTVLENPEAFSLTPAGRPQ
                     AIGPPQDDQERRRRTLASVASARLSAAAADSYWDTWPDVESNAGELLREYVSAPKAIM
                     EDLADNPIVAMTLLAHASLIASRNHPPYPAPATDREVILLEQREMMALLVGTHPAYAA
                     AFLGAPSFYAGLGLVSALARDGGLGDLLSDSVLTYRLVRSPASGRGGMPSTTRGSNDG
                     EDARRLTRHRIAGPPTGFIFFQDAWEEMDTRAALWPHPEFLGLVHNQSTARARACMLL
                     LARRCFAPEALQQLWHSLRPLEGPVAFQDYLRDFVKQAYTRGEELPRAEGLEVPRETP
                     SSYGTVTGRALRNLMPYGTPITGPKRGSGDTIPVSVFEAAVAAAFLGRPLTLFVSSQY
                     LFNLKTLGQVRVVAPLLYCDGHSEPFRSLVETISLNFLQDLDGYSESFEPEMSIFARQ
                     AVWLRELLTEARAAKPKEARPPTVAILANRKNIIWKCFTYRHNLPDVQFYFNAAGASR
                     WPTDVLNPSFYEHEDPPLPVGYQLPPNPRNVQELFSGFPPRVGHGLVSGDGFQSADNT
                     PASSDRLQQLGGGETDQGEKGSTTAESEASGPPSPQSPLLEKVAPGRPRDWLSPTSSP
                     RDVTVTPGLAAPITLPGPRLMARPYFGAETRASESPDRSPGSSPRPWPKDSLELLPQP
                     APQQPPSSPWASEQGPIVYTLSPHSTPSTASGSQKKHTIQIPGLVPSQKPSYPPSAPY
                     KPGQSTGGIAPTPSAASLTTFGLQPQDTQASSQDPPYGHSIMQREKKQQGGREEAAEI
                     RPSATRLPTAVGLRPRAPVVAAGAAASATPAFDPGEAPSGFPIPQAPALGSGLAAPAH
                     TPVGALAPRPQKTQAQRPQDAAALPTPTIKAVGARPVPKATGALAAGARPRGQPTAAP
                     PSAASPPRVSLPVRSRQQQSPAIPLPPMHSGSEPGARPEVRLSQYRHAGPQTYTVRKE
                     APPSAASQLPKMPKCKDSMYYPPSGSARYPAPFQALSFSQSVASPAPSSDQTTLLWNT
                     PSVVTQFLSIEDIIREVVTGGSTSGDLVVPSGSPSSLSTAAPEQDLRYSLTLSQASRV
                     LSRFVSQLRRKLERSTHRLIADLERLKFLYL"
misc_feature         62249
                     /note="BAM: BamH1 F/Q"
misc_feature         62430..62477
                     /note="Site III for EBNA-1 binding (Rawlins et al, 1985)"
misc_feature         66121
                     /note="BAM: BamH1 Q/U"
mRNA                 67477..67649
                     /note="Exon in EBNA-1 RNA (Speck and Strominger,1985) and
                     cDNA clone T4 (Bodescot et al, 1986)"
misc_feature         69410
                     /note="BAM: BamH1 U/P"
repeat_region        69684..69930
                     /note="5 x 51bp repeats"
repeat_region        70387..70521
                     /note="9 x 15bp repeat"
promoter             70750
                     /note="TATA: CATAAAA"
CDS                  complement(71520..75239)
                     /note="BOLF1 reading frame, 1 NXT/S analogous to VZV RF
                     21;  NCBI gi: 59096"
                     /codon_start=1
                     /translation="MASAMESDSSGGSGGADAQPPLAEVDGGLARVTRQLLLSGDDPA
                     ARLRALMPLELGIFGLGDLAQPVLVRDFLNTLTLMSGHAYPAAVLRHHAYYLLRAASF
                     SRRSFGLGHLEAALDVLASSLPPTTASPATDDPLDGSRLIAETRALAAAYRRIIEEGS
                     GEVLAVSGPTATFAFVEELVADTYLARWDAFPREGLSFYAFNAAKTTLGRWLVTVYAE
                     TNRYPWAAAGQGQPTAADIKAMAVELVEHSGGGAGGGEGEESGGGLFHRPESLSSVVA
                     SLPLARRRAVEILGVYAEASGGQTPPVAAVPVLAFDAARLRLLEPSGALFYDYVYEAL
                     LWDQTYGVPDSVIEAFLAGMAAEMEALAARVQEAAGSRASFSPAAIEQVATVLLSAGL
                     NETVAGDYAMMLASVPRVSRSRWRWLEATAALLESLSGFALHFFRLLPTASPTSRFAR
                     VARAAYLRAEAEAVDRRARRTSGPSTPAAAPAATAVGVGAAADPWDAVTPLRIFIVPP
                     PAAEYEQVAGDLSSELLRSLLWVRYSRLWQAPAPAPALPCKPPLLPGEQGRRQWTAAV
                     AAAPRTDVEAYCRSLRAGQTARADPAYVHSPFFPAAFIEFQIWPALRRVLSNELPKTR
```

```
                     SLAALRWLVSFGSDLALPSPELTRARRPLELIYATVWEIYDGAPPMPGESPQAVGLRP
                     LNLEGEGKAGDAGAEGAEDEEGGGPWGLSSHDAVLRIMDAVREVSGIISETISASERA
                     AEAPPLAWPTSLFSLLFTLRYSTTAESLGLATRRFLVSGETLSEDISRLTGAAWRLCS
                     RPLLYDAETGRVQIPLATEEEEEEAVVAVKEKSVSSSPRHYSTDLQTLKSVVEGIQDVC
                     RDAAARWALATADTATLRRRLLVPALRESRGIADHPLWAHTSEPLRPDLEELNERVEH
                     ALELGYSLTGALRRSVAYRFRDYTFARLFQPPAIDAERAEAIVRRDARPPPVFIPAPR
                     RLPQGGADTPPPLSMDDILYLGKSICKALVDVLDHHPAAPETTPIKTYTPAMDLNPEQ
                     ITVTPRSPSVLAAFARTARVQTHHLVPALTDDSPSPVGQTPPPFRILPAKKLAAILLG
                     NGRNASKRRASRDLSPPPHGRWRAVLDSSPFSFSSSDFSDQDEGEGGEADLRGVPGGG
                     GEGAYEEDRERPSDIDTAARAQKVETSCPRRRSPRTTPSPSRRASGGGGPDRGEAEAH
                     TYPPYLSAAAAASRVRPRTRRGATRRPPRPTAEDE"
promoter             complement(72192)
                     /note="TATA: TATTAAA before BPLF1"
misc_feature         73468
                     /note="BAM: BamH1 P/O"
promoter             75017
                     /note="TATA: TATTTAA BO-R1 late promoter before BORF1,
                     gives 3.9kb late RNA"
CDS                  75238..76332
                     /note="BORF1 late reading frame, 2 NXT/S homologous to VZV
                     RF20;  NCBI gi: 59095"
                     /codon_start=1
                     /translation="MKVQGSVDRRRLQRRIAGLLPPPARRLNISRGSEFTRDVRGLVE
                     EHAQASSLSAAAVWRAGLLAPGEVAVAGGGSGGGSFSWSGWRPPVFGDFLIHASSFNN
                     AEATGTPLFQFKQSDPFSGVDAVFTPLSLFIILMNHGRGVAARVEAGGGLTRMANLLYD
                     SPATLADLVPDFGRLVADRRFHNFITPVGPLVENIKSTYLNKITTVVHGPVVSKAIPR
                     STVKVTVPQEAFVDLDAWLSGGAGGGGGVCFVGGLGLQPCPADARLYVALTYEEAGPR
                     FTFFQSSRGHCQIMNILRIYYSPSIMHRYAVVQPLHIEELTFGAVACLGTFSATDGWR
                     RSAFNYRGSSLPVVEIDSFYSNVSDWEVIL"
promoter             complement(75322)
                     /note="TATA: TATTTAG before BOLF1"
promoter             75819
                     /note="TATA: TATAAAG"
misc_feature         75838
                     /note="polyA signal: AATAAA"
misc_feature         complement(76126)
                     /note="polyA signal: AATAAA"
promoter             76169
                     /note="TATA: TACATAT BO-R2 early promoter before BORF2,
                     gives 2.8kb RNA"
misc_feature         complement(76300)
                     /note="polyA signal: AATAAA"
CDS                  76407..78887
                     /note="BORF2 early reading frame, 2 NXT/S. Homology HSV
                     140K ribonucleotide reductase (Gibson et al, 1984) and RF
                     19 VZV;  NCBI gi: 59097"
                     /codon_start=1
                     /translation="MATTSHVEHELLSKLIDELKVKANSDPEADVLAGRLLHRLKAES
                     VTHTVAEYLEVFSDKFYDEEFFQMHRDELETRVSAFAQSPAYERIVSSGYLSALRYYD
                     TYLYVGRSGKQESVQHFYMRLAGFCASTTCLYAGLRAALQRARPEIESDMEVFDYYFE
                     HLTSQTVCCSTPFMRFAGVENSTLASCILTTPDLSSEWDVTQALYRHLGRYLFQRAGV
                     GVGVTGAGQDGKHISLLMRMINSHVEYHNYGCKRPVSVAAYMEPWHSQIFKFLETKLP
                     ENHERCPGIFTGLFVPELFFKLFRDTPWSDWYLFDPKDAGDLERLYGEEFEREYYRLV
                     TAGKFCGRVSIKSLMFSIVNCAVKAGSPFILLKEACNAHFWRDLQGEAMNAANLCAEV
                     LQPSRKSVATCNLANICLPRCLVNAPLAVRAQRADTQGDELLLALPRLSVTLPGEGAV
                     GDGFSLARLRDATQCATFVVACSILQGSPTYDSRDMASMGLGVQGLADVFADLGWQYT
                     DPPSRSLNKEIFEHMYFTALCTSSLIGLHTRKIFPGFKQSKYAGGWFHWHDWAGTDLS
                     IPREIWSRLSERIVRDGLFNSQFIAILMPTSGCAQVTGCSDAFYPFYANASTKVTNKEE
                     ALRPNRSFWRHVRLDDREALNLVGGRVSCLPEALRQRYLRFQTAFDYNQEDLIQMSRD
                     RAPFVDQSQSHSLFLREEDAARASTLANLLVRSYELGLKTIMYYCRIEKAADLGVMEC
```



```
                     KASAALSVPREEQNERSPAEQMPPRPMEPAQVAGPVDIMSKGPGEGPGGWCVPGGLEV
                     CYKYRQLFSEDDLLETDGFTERACESCQ"
misc_feature         77835
                     /note="BAM: Bam H1 O/a"
promoter             78804
                     /note="TATA: TATAAGT Ba-R1 early promoter before BaRF1,
                     gives 3.5kb RNA"
misc_feature         78883
                     /note="polyA signal: AATAAA, end of 3.9kb late RNA from
                     75017 and 2.8kb early RNA from 76169"
misc_feature         complement(78896)
                     /note="polyA signal: AATAAA"
CDS                  78900..79808
                     /note="BaRF1 early reading frame, homologous to HSV 38K
                     ribonucleotide reductase (Gibson et al, 1984) and RF 18
                     VZV;  NCBI gi: 59098"
                     /codon_start=1
                     /translation="MSKLLYVRDHEGFACLTVETHRNRWFAAHIVLTKDCGCLKLLNE
                     RDLEFYKFLFTFLAMAEKLVNFNIDELVTSFESHDIDHYYTEQKAMENVHGETYANIL
                     NMLFDGDRAAMNAYAEAIMADEALQAKISWLRDKVAAAVTLPEKILVFLLIEGIFFIS
                     SFYSIALLRVRGLMPGICLANNYISRDELLHTRAASLLYNSMTAKADRPRATWIQELF
                     RTAVEVETAFIEARGEGVTLVDVRAIKQFLEATADRILGDIGQAPLYGTPPPKDCPLT
                     YMTSIKQTNFFEQESSDYTMLVVDDL"
promoter             complement(79495)
                     /note="TATA: TATAACA"
misc_feature         79537
                     /note="BAM: Bam H1 a/M"
promoter             79840
                     /note="TATA: CATAAAT BM-R1 early promoter before BMRF1,
                     gives 2.5kb RNA"
CDS                  79899..81113
                     /note="BMRF1 early reading frame. Early antigen protein
                     recognised by R3 monoclonal (Pearson et al 1983; Cho et
                     al, 1985a);  NCBI gi: 59099"
                     /codon_start=1
                     /translation="METTQTLRFKTKALAVLSKCYDHAQTHLKGGVLQVNLLSVNYGG
                     PRLAAVANAGTAGLISFEVSPDAVAEWQNHQSPEEAPAAVSFRNLAYGRTCVLGKELF
                     GSAVEQASLQFYKRPQGGSRPEFVKLTMEYDDKVSKSHHTCALMPYMPPASDRLRNEQ
                     MIGQVLLMPKTASSLQKWARQQGSGGVKVTLNPDLYVTTYTSGEACLTLDYKPLSVGP
                     YEAFTGPVAKAQDVGAVEAHVVCSVAADSLAAALSLCRIPAVSVPILRFYRSGIIAVV
                     AGLLTSAGDLPLDLSVILFNHASEEAAASTASEPEDKSPRVQPLGTGLQQRPRHTVSP
                     SPSPPPPPRTPTWESPARPETPSPAIPSHSSNTALERPLAVQLARKRTSSEARQKQKH
                     PKKVKQAFNPLI"
promoter             80779
                     /note="TATA: TATTTAA BM-R2 late promoter before BMRF2"
misc_feature         complement(80782)
                     /note="polyA signal: AATAAA"
promoter             80832
                     /note="TATA: GATAAAA, possible promoter for 1.4kb late RNA
                     encoding BMRF2"
CDS                  81118..82191
                     /note="BMRF2 early reading frame;  NCBI gi: 59100"
                     /codon_start=1
                     /translation="MFSCKQHLSLGACVFCLGLLASTPFIWCFVFANLLSLEIFSPWQ
                     THVYRLGFPTACLMAVLWTLVPAKHAVRAVTPAIMLNIASALIFFSLRVYSTSTWVSA
                     PCLFLANLPLLCLWPRLAIEIVYICPAIHQRFFELGLLLACTIFALSVVSRALEVSAV
                     FMSPFFIFLALGSGSLAGARRNQIYTSGLERRRSIFCARGDHSVASLKETLHKCPWDL
                     LAISALTVLVVCVMIVLHVHAEVFFGLSRYLPLFLCGAMASGGLYLGHSSIIACVMAT
                     LCTLTSVVVYFLHETLGPLGKTVLFISIFVYYFSGVAALSAAMRYKLKKFVNGPLVHL
                     RVVYMCCFVFTFCEYLLVTFIKS"
```

```
promoter        81751
                /note="TATA: CATAAAT"
misc_feature    82180
                /note="polyA signal: ATTAAA, end of 3.5kb early RNA from
                78804, 2.5kb early RNA from 79840 and 1.4kb late RNA"
promoter        complement(82311)
                /note="TATA: CATAAAT"
repeat_region   82319..82461
                /note="2x71bp repeats"
CDS             complement(82743..84122)
                /note="BMLF1 early reading frame. Diffuse early antigen
                (Cho et al, 1985b). Also homologous to RF 4 VZV and IE63
                of HSV.(BSLF2 + BMLF1) is also called EB2
                (Chevallier-Greco et al, 1986). General transactivator of
                transcription (Lieberman et al, 1986). NCBI gi: 583888"
                /codon_start=1
                /translation="XSHILELEAVSDTNTDCDLDPMEGSEEHSTDGEISSSEEEDEDP
                TPAHAIPARPSSVVITPTSASFVIPRKKWDLQDKTVTLHRSPLCRDEDEKEETGNSSY
                TRGHKRRRGEVHGCTDESYGKRRHLPPGARAPRAPRAPRVPRAPRSPRAPRSNRATRG
                PRSESRGAGRSTRKQARQERSQRPLPNKPWFDMSLVKPVSKITFVTLPSPLASLTLEP
                IQDPFLQSMLAVAAHPEIGAWQKVQPRHELRRSYKTLREFFTKSTNKDTWLDARMQAI
                QNAGLCTLVAMLEETIFWLQEITYHGDLPLAPAEDILLACAMSLSKVILTKLKELAPC
                FLPNTRDYNFVKQLFYITCATARQNKVVETLSSSYVKQPLCLLAAYAAVAPAYINANC
                RRRHDEVEFLGHYIKNYNPGTLSSLLTEAVETHTRDCRSASCSRLVRAILSPGTGSLG
                LFFVPGLNQ"
misc_feature    complement(82747)
                /note="polyA signal: AATAAA"
repeat_region   83640..83729
                /note="10x9bp repeats"
misc_feature    complement(84122)
                /note="ACCEPT: CTCCCCTCTGCAG acceptor in spliced form of
                BMLF1 RNA"
misc_feature    complement(84227)
                /note="DONOR: CAGGTAAGA donor in spliced form of BMLF1
                RNA"
CDS             complement(84229..84288)
                /note="BSLF2 early reading frame in 5' exon of spliced RNA
                encoding BMLF1; NCBI gi: 59102"
                /codon_start=1
                /translation="MVPSQRLSRTSSISSNEDPA"
misc_feature    84233
                /note="BAM: Bam H1 M/S"
CDS             complement(84257..86881)
                /note="BSLF1 reading frame, homologous to RF 6 VZV; NCBI
                gi: 59103"
                /codon_start=1
                /translation="MSAPVVIKALVASNTDIAEAILDAILSRPDEGFRLFCLCHNASP
                LHHVAGSLVELQLHLPKKRLTSQSRCGLVLTLHLPAEEAFPFLRGLTPLTADRLSTYL
                DRAGALRSLTPLVELLTLSAKKQPQGDARGRVAWLRPKIVGCLRRIYRVNISARWFIS
                TFGSHEAQFVLVTAAYYFWGIPCTIETLAHLTELFTSESGQSLAAVTSLAELGEVFGS
                SAWAEQTEAFAHFAHEKLRRDSREIRAVARTIDAYRGRLPLASADLVRYVYLAHAQCF
                NEGTFKRYSQLTSMGEIGCLPSGGVVLPSLLDRGFAEHMRTYFTRETYLAEHVRVQQL
                KIRMEPPAPYTWDPDPDDGLMRAWAGLSVDVARELVELARWHADEGPTYPPTLQGFLC
                LAGQATCRGQWNPKEQFLPPTVLRRVQRLPVFLCHFADRHYFVMTAADPFSSHLAEVV
                STPTNCRLPDTCLTRALSYTPVYYSQNSLSEQLFVSRHEYFNPRLPVCNLVLDLDLKI
                KGAPWSLEEIYDLCRTVRREVLRLMRRLGPVSRAHPVYFFKSACPPADPDNMEDVLPF
                CICTGKLGFRVITPLPRGHAIVGTSAVQGFVSVLQKLMGLTACLRRMRHKIKEIGAPL
                FDSGVYHAGRCIRLPHTYKVDRGGGLSRQLRLFVCHPEEEDKHSYVKNALNIQNLLHH
                SLHVGWPAPKTFCYHIADDGRDYLIQRTRETLPPTVENVCAMIEGHLGLDLVAWVSSC
                IWPSLMSTLATAVPEDKFPQFLHVTFEQTGPNLVQVCHARGRNFACLRHTHRASSKNV
```



```
                RVFLVLYYTSQAITVTFMSQCFAGRCGANQPTAHFSISVPASRIINRAEASQDSTTSQ
                LARRRDRQDGSFSETLPN"
promoter        complement(84356)
                /note="TATA: CATAAAT before BSLF2 and BMLF1. Two RNAs
                start here; one is spliced and the other is unspliced,
                both traverse BMLF1."
promoter        86882
                /note="TATA: TATTTAA BS-R1 late promoter before BSRF1"
CDS             86924..87580
                /note="BSRF1 reading frame; NCBI gi: 59104"
                /codon_start=1
                /translation="MAFYLPDWSCCGLWLFGRPRNRYSQLPEEPETFECPDRWRAEID
                LGLPPGVQVGDLLRNEQTMGSLRQVYLLAVQANSITDHLKRFDAVRVPESCRGVVEAQ
                VAKLEAVRSVIWNTMISLAVSGIEMDENGLKALLDKQAGDSLAIMEMEKVATALKMDE
                TGAWAQEISAVVSSVTAPSASAPFINSAFEPEVPTPVLAPPPVVRQPEHSGPTELALT
                "
misc_feature    complement(87134)
                /note="polyA signal: AATAAA"
misc_feature    87599
                /note="polyA signal: AATAAA"
misc_feature    complement(87613)
                /note="polyA signal: AATAAA, end 1.0kb early RNA from
                BLL3"
CDS             complement(87638..88474)
                /note="BLLF3 early reading frame (BLLF2 in Baer et al,
                1984). Homologous to RF 8 VZV and dUTPase HSV. NCBI gi:
                59105"
                /codon_start=1
                /translation="MEACPHIRYAFQNDKLLLQQASVGRLTLVNKTTILLRPMKTTTV
                DLGLYARPPEGHGLMLWGSTSRPVTSHVGIIDPGYTGELRLILQNQRRYNSTLRPSEL
                KIHLAAFRYATPQMEEDKGPINHPQYPGDVGLDVSLPKDLALFPHQTVSVTLTVPPPS
                IPHHRPTIFGRSGLAMQGILVKPCRWRRGGVDVSLTNFSDQTVFLNKYRRFCQLVYLH
                KHHLTSFYSPHSDAGVLGPRSLFRWASCTFEEVPSLAMGDSGLSEALEGRQGRGFGSS
                GQ"
misc_feature    87650
                /note="BAM: Bam H1 S/L"
promoter        88507
                /note="TATA: TATATAT BL-R1 late promoter before BLRF1,
                gives 1.0kb late RNA"
promoter        88511
                /note="TATA: TATAAGA"
promoter        complement(88514)
                /note="TATA: TATATAT BL-L3 early promoter before BLLF3,
                gives 1.0kb early RNA"
CDS             88547..88855
                /note="BLRF1 late reading frame; NCBI gi: 59106"
                /codon_start=1
                /translation="MGKVLRKPFAKAVPLLFLAATWLLTGVLPAGASSPTNAAAASLT
                EAQDQFYSYTCNADTFSPSLTSFASIWALLTLVLVIIASAIYLMYVCFNKFVNTLLTD
                "
promoter        88863
                /note="TATA: TATTTAA BL-R2 late promoter before BLRF2,
                gives 0.6kb late RNA"
CDS             88925..89413
                /note="BLRF2 late reading frame, 2 NXS/T; NCBI gi: 59107"
                /codon_start=1
                /translation="MSAPRKVRLPSVKAVDMSMEDMAARLARLESENKALKQQVLRGG
                ACASSTSVPSAPVPPPEPLTARQREVMITQATGRLASQAMKKIEDKVRKSVDGVTTRN
                EMENILQNLTLRIQVSMLGAKGQPSPGEGTRPRESNDPNATRRARSRSRGREAKKVQI
                SD"
```

misc_feature 89412
/note="polyA signal: AATAAA, end of 1.0kb and 0.6kb late RNAs"
misc_feature complement(89425)
/note="polyA signal: AATAAA, end of 0.7kb early, 2.2kb late and 2.8kb late RNA"
CDS complement(89430..92153)
/note="BLLF1a, late reading frame, gp350 membrane antigen, 36 NXT/S (Hummel et al, 1984; Biggin et al, 1984; Beisel et al, 1985); NCBI gi: 59110"
/codon_start=1
/translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
VSGLGTAPKTLIIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
CDS complement(89430..92153)
/note="BLLF1b, late reading frame gp220 membrane antigen, spliced form of BLLF1a (Hummel et al, 1984; Biggin et al, 1984; Beisel et al, 1985); NCBI gi: 59109"
/codon_start=1
/translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
VSGLGTAPKTLIIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
promoter complement(89434)
/note="TATA: TATAAAG"
CDS complement(89567..90013)
/note="BLLF2 early reading frame (BLLF3 in Baer et al, 1984); NCBI gi: 59108"
/codon_start=1
/translation="MCPPVRQHPAQAPPAKRQALETVPHPQNRGRLMSPKARPPKMQR
RPRPPVAKRRRFPRSPQQVERPILPPVESTPQDMEPGQVQSPPQITAVIQLRQDRDTM
RPPIYLPALLANCGPAGLLRAHRLPQPKPPCQSRQRPSPDSQTSPC"
promoter complement(90051)
/note="TATA: TATAACA BL-L2 early promoter before BLLF2, gives 0.7kb early RNA"

intron complement(90062..90652)
/note="intervening sequence in gp220 gene"
repeat_region 90177..90639
/note="21 copies of 21bp approximate repeat"
promoter complement(92192)
/note="TATA: TATTAAA BL-L1 late promoter before BLLF1a,b. Gives 2.8 and 2.2kb late RNAs"
mRNA 92238..92581
/note="Exon in Bodescot et al (1986) RNA (spliced from 20763 to 92670)"
CDS 92243..92581
/note="BLRF3 reading frame; NCBI gi: 59111"
/codon_start=1
/translation="MDKDRPGPPALDDNMEEEVPSTSVVQEQVSAGDWENVLIELSDS
SSEKEAEDAHLEPAQKGTKRKRVDHDAGGSAPARPMLPPQPDLPGREAILRRFPLDLR
TLLQAIGAAAT"
CDS 92670..95165
/note="BERF1 frame, homology with BERF2b and BERF4. A fusion of BLRF3 with BERF1 encodes EBNA-3A, latent cycle gene. (Hennessy et al, 1986, Joab et al, 1987); NCBI gi: 583889"
/codon_start=1
/translation="XIDTRAIDQFFGSQISNTEMYIMYAMAIRQAIRDRRRNPASRRD
QAKWRLQTLAAGWPMGYQAYSSWMYSYTDHQTTPTFVHLQATLGCTGGRRCHVTFSAG
TFKLPRCTPGDRQWLYVQSSVGNIVQSCNPRYSIFFDYMAIHRSLTKIWEEVLTPDQR
VSFMEFLGFLQRTDLSYIKSFVSDALGTTSIQTPWIDDNPSTETAQAWNAGFLRGRAY
GIDLLRTEGEHVEGATGETREESEDTESDGDDEDLPCIVSRGGPKVKRPPIFIRRLHR
LLLMRAGKRTEQGKEVLEKARGSTYGTPRPPVPKPRPEVPQSDETATSHGSAQVPEPP
TIHLAAQGMAYPLHEQHGMAPCPVAQAPPTPLPPVSPGDQLPGVFSDGRVACAPVPAP
AGPIVRPWEPSLTQAAGQAFAPVRPQHMPVEPVPVPTVALERPVYPKPVRPAPPKIAM
QGPGETSGIRRARERWRPAPWTPNPPRSPSQMSVRDRLARLRAEAQVKQASVEVQPPQ
LTQVSPQQPMEGPLVPEQQMFPGAPFSQVADVVRAPGVPAMQPQYFDLPLIQPISQGA
PVAPLRASMGPVPPVPATQPQYFDIPLTEPINQGASAAHFLPQQPMEGPLVPEQWMFP
GAALSQSVRPGVAQSQYFDLPLTQPINHGAPAAHFLHQPPMEGPWVPEQWMFQGAPPS
QGTDVVQHQLDALGYTLHGLNHPGVPVSPAVNQYHLSQAAFGLPIDEDESGEGSDTSE
PCEALDLSIHGRPCPQAPEWPVQEEGGQDATEVLDLSIHGRPRPRTPEWPVQGEGGQN
VTGPETRRVVVSAVVHMCQDDEFPDLQDPPDEA"
mRNA 92670..95248
/note="Exon in (Bodescot et al, 1986) RNA from 92581, to 3' end"
misc_feature 92703
/note="BAM: Bam H1 L/E"
promoter complement(93161)
/note="TATA: CATAAAT"
promoter 93479
/note="TATA: TATAAGA"
promoter complement(93482)
/note="TATA: TATAAAT"
repeat_region 94208..94277
/note="repeat type A"
repeat_region 94281..94306
/note="repeat type B"
repeat_region 94307..94381
/note="repeat type C"
repeat_region 94386..94411
/note="repeat type B"
repeat_region 94412..94489
/note="repeat type C"
repeat_region 94490..94560
/note="repeat type A"

```
repeat_region   94571..94648
                /note="repeat type C"
repeat_region   94649..94719
                /note="repeat type A"
repeat_region   94896..94982
                /note="repeat type D"
repeat_region   94983..95069
                /note="repeat type D"
misc_feature    95221
                /note="polyA signal: AATAAA"
misc_feature    complement(95272)
                /note="polyA signal: AATAAA"
CDS             95353..95724
                /note="BERF2a reading frame;  NCBI gi: 59113"
                /codon_start=1
                /translation="MKKAWLSRAQQADAGGASGSEDPPDYGDQGNVTQVGSEPISPEI
                GPFELSAASEDDPQSGPVEENLDAAAREEEEPHEQEHNGGDDPLDVHTRQPRFVDVNP
                TQAPVIQLVHAVYDSMLVRGT"
CDS             95725..98247
                /note="BERF2b frame, homology with BERF1 and BERF4. BERF2a
                and BERF2b are spliced together to make EBNA3B (EBNA4A)
                latent protein.  NCBI gi: 583890"
                /codon_start=1
                /translation="XISRCFAWIFWPVLIDCHWFQQSDLRPLGSLFLEQNLNIEEFIW
                MCMTVRHRCQAIRKKPLPIVKQRRWKLLSSCRSWRMGYRTHNLKVNSFESGGDNVHPV
                LVTATLGCDEGTRHATTYSAGIVQIPRISDQNQKIETAFLMARRARSLSAERYTLFFD
                LVSSGNTLYAIWIGLGTKNRVSFIEFVGWLCKKDHTHIREWFRQCTGRPKAAKPWLRA
                HPVAIPYDDPLTNEEIDLAYARGQAMNIEAPRLPDDPIIVEDDDESEEIEAESDEEED
                KSGMESLKNIPQTLPYNPTVYGRPAVFDRKSDAKSTKKCRAIVTDFSVIKAIEEEHRK
                KKAARTEQPRATPESQAPTVVLQRPPTQQEPGPVGPLSVQARLEPWQPLPGPQVTAVL
                LHEESMQGVQVHGSMLDLLEKDDEVMEQRVMATLLPPVPQQPRAGRRGPCVFTGDLGI
                ESDEPASTEPVHDQLLPAPGPDPLEIQPLTSPTTSQLSSSAPSCAQTPWPVVQPSQTP
                DDPTKQSRPPETAAPRQWPMPLRPIPMRPLRMQPIPFNHPVGPTPHQTPQVEITPYKP
                TWAQIGHIPYQPTPTGPATMLLRQWAPATMQTPPRAPTPMSPPEVPPVPRQRPRGAPT
                PTPPPQVPPVPRQRPRGAPTPTPPPQVLPTPMQLALRAPAGQQGPTKQILRQLLTGGV
                KKGRPSLKLQAALERQAAAGWQPSPGSGTSDKIVQAPIFYPPVLQPIQVMGQGGSPTA
                MAASAVTQAPTEYTRERRGVGPMPPTDIPPSKRAKIEAYTEPEMPHGGASHSPVVILE
                NVGQGQQQTLECGGTAKQERDMLGLGDIAVSSPSSSETSNDE"
misc_feature    complement(95819)
                /note="polyA signal: AATAAA"
promoter        complement(95853)
                /note="TATA: TATAAAT"
misc_feature    complement(96276)
                /note="polyA signal: AATAAA"
repeat_region   97522..97698
                /note="3x60bp repeat"
CDS             98323..98769
                /note="BERF3 reading frame;  NCBI gi: 59115"
                /codon_start=1
                /translation="MTILDEVEILHIFRPTMESFEGQGDSRQSPDNERGDNVQTTGEH
                DQDPGPGPPSSGASERLVPEESYSRDQQPWGQSRGDENRGWMQRIRRRRRRRAALSGH
                LLDTEDNVPPWLPPHDITPYTARNIRDAACRAVKVSMPLTGFMGAI"
mRNA            98364..98730
                /note="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"
misc_feature    98731
                /note="DONOR: AAGGTGAGT donor"
CDS             98805..101423
                /note="BERF4 frame, homology with BERF1 and BERF2b. BERF3
                and BERF4 are spliced together to make the EBNA3C (EBNA
                4B) latent protein.  NCBI gi: 583891"
```



```
                /codon_start=1
                /translation="XSHLQALSNLILDSGLDTQHILCFVMAARQRLQDIRRGPLVAEG
                GVGWRHWLLTSPSQSWPMGYRTATLRTLTPVPNRVGADSIMLTATFGCQNAARTLNTF
                SATVWTPPHAGPREQERYAREAEVRFLRGKWQRRYRRIYDLIELCGSLHHIWQNLLQT
                EENLLDFVRFMGVMSSCNNPAVNYWFHKTIGNFKPYYPWNAPPNENPYHARRGIKEHV
                IQNAFRKAQIQGLSMLATGGEPRGDATSETSSDEDTGRQGSDVELESSDDELPYIDPN
                MEPVQQRPVMFVSRVPAKKPRKLPWPTPKTHPVKRTNVKTSDRSDKAEAQSTPERPGP
                SEQSSVTVEPAHPTPVEMPMVILHQPPPVPKPVPVKPTPPPSRRRRGACVVYDDDVIE
                VIDVETTEDSSSVSQPNKPHRKHQDGFQRSGRRQKRAAPPTVSPSDTGPPAVGPPAAG
                PPAAGPPAAGPPAAGPPAAGPPAAGPRILAPLSAGPPAAGPHIVTPPSARPRIMAPPV
                VRMFMRERQLPQSTGRKPQCFWEMRAGREITQMQQEPSSHLQSATQPTTPRPSWAPSV
                CALSVMDAGKAQPIESSHLSSMSPTQPISHEEQPRYEDPDAPLDLSLHPDVAAQPAPQ
                APYQGYQEPPAPQAPYQGYQEPPPPQAPYQGYQEPPAHGLQSSSYPGYAGPWTPRSQH
                PCYRHPWAPWSQDPVHGHTQGPWDPRAPHLPPQWDGSAGHGQDQVSQFPHLQSETGPP
                RLQLSLVPLVSSSAPSWSSPQPRAPIRPIPTRFPPPPPMPLQDSMAVGCDSSGTACPSM
                PFASDYSQGAFTPLDINATTPKRPRVEESSHGPARCSQATAEAQEILSDNSEISVFPK
                DAKQTDYDASTESELD"
mRNA            98805..99050
                /note="Exon in T4 cDNA (Bodescot et al 1986). 99050 is not
                the end of the RNA."
misc_feature    99126..102118
                /note="DEL: Deletion in Raji"
promoter        99443
                /note="TATA: CATAAAA"
misc_feature    100104
                /note="DONOR: ACCGTGAGT possible donor before repeat."
repeat_region   100122..100304
                /note="10 x 15bp repeat"
misc_feature    100528
                /note="DONOR: CTGGTAAGG possible donor"
misc_feature    100613
                /note="BAM: Bam H1 E/e1"
repeat_region   100665..100781
                /note="3x39bp repeat"
promoter        complement(100860)
                /note="TATA: TATAACA"
misc_feature    100919
                /note="BAM: Bam H1 e1/e2"
misc_feature    101426
                /note="BAM: Bam H1 e2/e3"
CDS             complement(101445..102116)
                /note="BZLF2 reading frame 3x NXT/S. 2.5kb late RNA
                traverses BZLF2, ends unknown.  NCBI gi: 59117"
                /codon_start=1
                /translation="MVSFKQVRVPLFTAIALVIVLLLAYFLPPRVRGGGRVAAAAITW
                VPKPNVEVWPVDPPPPVNFNKTAEQEYGDKEVKLPHWTPTLHTFQVPQNYTKANCTYC
                NTREYTFSYKGCCFYFTKKKHTWNGCFQACAELYPCTYFYGPTPDILPVVTRNLNAIE
                SLWVGVYRVGEGNWTSLDGGTFKVYQIFGSHCTYVSKFSTVPVSHHECSFLKPCLCVS
                QRSNS"
promoter        101690
                /note="TATA: CATAAAA"
misc_feature    101765
                /note="polyA signal: AATAAA"
promoter        complement(101786)
                /note="TATA: TATAAAG"
misc_feature    101947
                /note="BAM: Bam H1 e3/Z"
misc_feature    complement(102098)
                /note="DONOR: CAGGTGAGG possible donor"
mRNA            complement(102126..102341)
```

```
                /note="3' terminal exon of 0.9kb and 2.8kb early RNAs"
promoter        102153
                /note="TATA: TATTAAT"
misc_feature    complement(102156)
                /note="polyA signal: AATAAA 3' end of 0.9kb and 2.8kb RNAs
                encoding BZLF1 and BRLF1"
promoter        complement(102160)
                /note="TATA: TATTAAT"
CDS             complement(102213..103155)
                /note="BZLF1 reading frame, modified from Baer et al,
                1984. Has two splices within frame. 2xNXT/S. Immediate
                early gene which disrupts latency (Countryman and Miller,
                1985), called EB1 by Chevallier-Greco et al, 1986 and
                ZEBRA by Miller."
                /codon_start=1
promoter        complement(102380)
                /note="TATA: CATAAAT"
promoter        102415
                /note="TATA:TATATAC"
promoter        complement(102420)
                /note="TATA: TATATAC"
mRNA            complement(102426..102530)
                /note="Exon of 0.9kb and 2.8kb early RNAs"
misc_feature    complement(102504)
                /note="polyA signal: AATAAA, apparently not functional"
repeat_region   102581..102652
                /note="semi-repetitive sequence, homologous to human c-fos
                3' sequence"
mRNA            complement(102655..103194)
                /note="First exon of 0.9kb early RNA encoding BZLF1"
misc_feature    complement(102918)
                /note="splice acceptor used in RZ fusion gene (Sargeant)"
promoter        complement(103231)
                /note="TATA: TTTAAA of BZL1 immediate early promoter gives
                0.9kb RNA"
misc_feature    complement(103256..103311)
                /note="Upstream of BZL1, homology to 106243 to 106188"
CDS             complement(103366..105183)
                /note="BRLF1 reading frame, (immediate?) early gene, acts
                as transcription activator.  NCBI gi: 59119"
                /codon_start=1
                /translation="MRPKKDGLEDFLRLTPEIKKQLGSLVSDYCNVLNKEFTAGSVEI
                TLRSYKICKAFINEAKAHGREWGGLMATLNICNFWAILRNNRVRRRAENAGNDACSIA
                CPIVMRYVLDHLIVVTDRFFIQAPSNRVMIPATIGTAMYKLLKHSRVRAYTYSKVLGV
                DRAAIMASGKQVVEHLNRMEKEGLLSSKFKAFCKWVFTYPVLEEMFQTMVSSKTGHLT
                DDVKDVRALIKTLPRASYSSHAGQRSYVSGVLPACLLSTKSKAVETPILVSGADRMDE
                ELMGNDGGASHTEARYSESGQFHAFTDELESLPSPTMPLKPGAQSADCGDSSSSSSDS
                GNSDTEQSEREEARAEAPRLRAPKSRRTSRPNRGQTPCPSNAAEPEQPWIAAVHQESD
                ERPIFPHPSKPTFLPPVKRKKGLRDSREGMFLPKPEAGSAISDVFEGREVCQPKRIRP
                FHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDP
                DEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNL
                DSPLTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF"
misc_feature    complement(103453..103462)
                /note="TAATGAAATC sequence"
misc_feature    103741
                /note="BAM: Bam H1 Z/g"
misc_feature    103816
                /note="BAM: Bam H1 g/R"
mRNA            complement(104926..105185)
                /note="exon in RZ fusion gene (Sargeant)"
```



```
mRNA            complement(104927..104989)
                /note="BRLF2 poss. small 5' exon"
promoter        105016
                /note="TATA: TATAAAT before BRRF1, possible promoter for
                1.1 kb early RNA encoding BRRF1"
promoter        complement(105019)
                /note="TATA: TATAAAT before BRLF2"
CDS             105182..106114
                /note="BRRF1 early reading frame;  NCBI gi: 59118"
                /codon_start=1
                /translation="MASSNRGNARPLKSFLHELYLKHYPEVGDVVHLLNTIGVDCDLP
                PSHPLLTAQRGLFLARVLQAVQQHKLLEDTIVPKILKKLAYFLELLSYYSPKDEQRDI
                AEVLDHLKTNRDLGLDDRLWALIRKLRQDRHHASVNVLMPGSDYTAVSLQYYDGISIG
                MRKVIADVCRSGYASMPSMTATHNLSHQLLMASGPSEEPCAWRGFFNQVLLWTVALCK
                FRRCIYYNYIQGSIATISQLLHLEIKALCSWIISQDGMRLFQHSRPLLTLWESVAANQ
                EVTDAITLPDCAEYIDLLKHTKHVLENCSAMQYK"
misc_feature    complement(105185)
                /note="ACCEPT: splice acceptor in 2.8kb early RNA encoding
                BRLF1 and RZ fusion gene (Sargeant)"
promoter        105213
                /note="TATA: CATTAAA"
misc_feature    106110
                /note="polyA signal: AATAAA, 3' end of early 1.1kb RNA
                encoding BRRF1"
misc_feature    complement(106125)
                /note="DONOR: CAGGTAAGA possible donor"
misc_feature    complement(106188..106243)
                /note="Homology to upstream region of BZL1"
promoter        complement(106213)
                /note="TATA: CATAAAA"
promoter        106243
                /note="TATA: TATAAAA before BRRF2, possible promoter for
                1.8 kb RNA encoding BRRF2"
CDS             106302..107915
                /note="BRRF2 reading frame;  NCBI gi: 59120"
                /codon_start=1
                /translation="MSGQQRGSVILVPEHLAGALTKLMSDFITGQDVTLSGGNIAVKI
                RDAINQTPGGGDVAILSSLFALWNALPTSGRQSSRDDLIPAAVQALTTAHNLCLGVIP
                GETSHKDTPESLLRAIVTGLQKLWVDSCGCPECLQCLKGLKAIKPGLYEIPRIIPHTK
                QCSPVNLLNMLVHKLVALRGHVQLAYDARVLTPDFHEIPDLDDSDAVFARTLLAALFH
                LNMFFILKDYITQDSMSLKQALSGHWMSATGNPLPAAPETLRDYLEAFRNSDNHFYLP
                TTGPLNTFQFPEELLGRVVVIDSSLCAASHVQDVITHGVGAGVPRPRFSALPPAPSRE
                PQQTCSQLTSRGNESSRRNLGQPGGTSPAVPPVCPIVSLTASGAKQNRGGMGSLHLAK
                PEETSPAVSPVCPIASPAASRSKQHCGVTGSSQAAPSFSSVAPVASLSGDLEEEEEGS
                RESPSLPSSKKGDEEFEAWLEAQDANLEDVQREFSGLRVIGDEDEDGSEDGEFSDLDL
                SDSDHEGDEGGGAVGGGRSLHSLYSLSVV"
promoter        complement(106385)
                /note="TATA: GATAAAA"
misc_feature    complement(106973)
                /note="polyA signal: AATAAA"
promoter        complement(107124)
                /note="TATA: GATAAAA"
misc_feature    107457
                /note="BAM: Bam H1 R/f"
misc_feature    107565
                /note="BAM: Bam H1 f/K"
misc_feature    107914
                /note="polyA signal: AATAAA, 3' end of 1.8kb RNA encoding
                BBRF2"
misc_feature    107942
```

```
                     /note="ACCEPT: splice acceptor for EBNA-1 RNA (from
                     98730)"
     CDS             107950..109875
                     /note="BKRF1 encodes EBNA-1 protein, latent cycle gene.
                     NCBI gi: 59121"
                     /codon_start=1
                     /translation="MSDEGPGTGPGNGLGEKGDTSGPEGSGGSGPQRRGGDNHGRGRG
                     RGRGRGGRPGAPGGSGSGPRHRDGVRRPQKRPSCIGCKGTHGGTGAGAGAGGAGAGG
                     AGAGGAGAGGGAGGAGGAGGAGACGGGAGAGGGAGGAGGAGAGGGAGAGGGAGGAGAG
                     GGAGGAGGAGAGGGAGAGGGAGGAGAGGGAGGAGGAGAGGGAGGAGAGGAGAGGA
                     GAGGGAGGAGGAGAGGAGAGGAGAGGAGAGGAGGAGAGGAGGAGAGGAGAGGGAG
                     GAGAGGGAGGAGAGGAGGAGAGGAGGAGAGGAGGAGAGGGAGAGGAGAGGGGRGRGGS
                     GGRGRGGSGGRGRGGSGGRRGRGRERARGGSRERARGRGRGRGEKRPRSPSSQSSSSG
                     SPPRRPPPGRRPFFHPVGEADYFEYHQEGGPDGEPDVPPGAIEQGPADDPGEGPSTGP
                     RGQGDGGRRKKGGWFGKHRGQGGSNPKFENIAEGLRALLARSHVERTTDEGTWVAGVF
                     VYGGSKTSLYNLRRGTALAIPQCRLTPLSRLPFGMAPGPGPQPGPLRESIVCYFMVFL
                     QTHIFAEVLKDAIKDLVMTKPAPTCNIRVTVCSFDDGVDLPPWFPPMVEGAAAEGDDG
                     DDGDEGGDGDEGEEGQE"
     repeat_region   108217..108924
                     /note="EBNA triplet repeat GGA,GCA,GGG."
     misc_feature    109856
                     /note="DONOR: AGGGTGAGG possible donor at end BKRF1"
     promoter        109905
                     /note="TATA: TATTAAA before BKRF2, possible start for
                     2.3kb late RNA"
     misc_feature    109906
                     /note="polyA signal: ATTAAA"
     misc_feature    109937
                     /note="polyA signal: AATAAA 3' end of EBNA-1 RNA"
     CDS             109958..110371
                     /note="BKRF2 reading frame;  NCBI gi: 59122"
                     /codon_start=1
                     /translation="MRAVGVFLAICLVTIFVLPTWGNWAYPCCHVTQLRAQHLLALEN
                     ISDIYLVSNQTCDGFSLASLNSPKNGSNQLVISRCANGLNVVSFFISILKRSSSALTG
                     HLRELLTTLETLYGSFSVEDLFGANLNRYAWHRGG"
     misc_feature    110271
                     /note="DONOR: TCCGTGAGT possible donor at end of BKRF2"
     CDS             110275..111120
                     /note="BKRF3 reading frame, homologous to RF 59 VZV;  NCBI
                     gi: 583892"
                     /codon_start=1
                     /translation="XVNHPGDSLRFILSGRPVWCQLKQIRMASRGLDLWLDEHVWKRK
                     QEIGVKGENLLLPDLWLDFLQLSPIFQRKLAAVIACVRRLRTQATVYPEEDMCMAWAR
                     FCDPSDIKVVILGQDPYHGGQANGLAFSVAYGFPVPPSLRNIYAELHRSLPEFSPPDH
                     GCLDAWASQGVLLLNTILTVQKGKPGSHADIGWAWFTDHVISLLSERLKACVFMLWGA
                     KAGDKASLINSKKHLVLTSQHPSPLAQNSTRKSAQQKFLGNNHFVLANNFLREKGLGE
                     IDWRL"
     misc_feature    111098
                     /note="DONOR: TCGGTGAGA possible donor at end BKRF3"
     CDS             111107..111787
                     /note="BKRF4 reading frame, contains complex repetitive
                     sequence;  NCBI gi: 583893"
                     /codon_start=1
                     /translation="XGGCRGVITMAMFLKSRGVRSCRDRRLLSDEEEETSQSSSYTLG
                     SQASQSIQEEDVSDTDESDYSDEDEEIDLEEEYPSDEDPSEGSDSDPSWHPSDSDESD
                     YSESDEDEATPGSQASRSSRVSPSTQQSSGLTPTPSFSRPRTRAPPRPPAPAPVRGRA
                     SAPPRPPAPVQQSTKDKGPHRPTRPVLRGPAPRRPPPPSSPNTYNKHMMETTPPIKGN
                     NNYNWPWL"
     misc_feature    111272
                     /note="DONOR: GACGTGAGT poss.donor before rpt.seq. in
```



```
                     BKRF4"
     misc_feature    111719
                     /note="polyA signal: AATAAA"
     misc_feature    111787
                     /note="polyA signal: AATAAA : currently unknown which is
                     3' end of the 2.3kb late and 1.1kb early RNAs"
     CDS             complement(111830..114259)
                     /note="BBLF4 early reading frame, very good homology to
                     RF55 VZV;  NCBI gi: 59126"
                     /codon_start=1
                     /translation="MAEEPRAPEALSSTFMLNMTSDASVRRIVRRIGTLARRRVQQLP
                     DMETFSPEFDPELSEPPFLPFSAYVITGTAGAGKSTSVSCLHHTMDCLVTGATTVAAQ
                     NLSQTLRAYCPTVYSAFGFKSRHINMTQRVSSHGRSTDAALEELQRRDLAKYWPVLSD
                     IAAEFRRTKPRGLYSGVSGPAFEVLRDMHQGQLWTTNVIVVDEAGTLSVHILTAVVFC
                     YWFFNAWLRTPLYRRGRIPCIVCVGSPTQTDAFQSSFSHETQVNKIRECDNILTFLVG
                     NPRAATYVDVARNWALFINNKRCTDVQFGHLMKTLEYGLELSPDILAYVDRFVVPRAA
                     IMDPAQYVGWTRLFLSHAEVKTFLTTLHATLKTAGQGRAARGTGGDGGGVTMFTCPVE
                     CEVFLDPLAQYKTLVGLPGLTAHTWLQKNYARLGNYSQFADQDMVPVGTEQDEERVKV
                     TYNVTYVKHSSVSVNCKTKKSICGYTGTFGDFMDTLEADSFVEAHGHEQPEYVYSFLA
                     RLIYGGIYAFSHGGHSLCENGEYVAELGAVPLPGRTWDPEVTAGMELGELPLEVAWDG
                     ERSPAAVFYARVLAPPAANSAPLCSLLNIYNDLRAYFRQCLDVAVRYGGREFRDLPFC
                     TFTNNMLIRDNIEFTSDEPLLHGLLDYASTTENYTLLGYTHLNVFFGIRGKQQPQDAG
                     SSRMPRLMVKDEAGFVCCLEHNTNKLYETIEDKSLNLCSIRDYGISSKLAMTIAKAQG
                     LSLNKVAICFGSHRNIKPGHVYVALSRARHSNCVVMDRNPLSEMITGEGNPASGYIVD
                     ALKNSRALLVY"
     misc_feature    complement(111830)
                     /note="polyA signal: AATAAA"
     promoter        112471
                     /note="TATA: TATATAT"
     promoter        complement(112476)
                     /note="TATA: TATATAA"
     misc_feature    112620
                     /note="BAM: Bam H1 K/B"
     promoter        113876
                     /note="TATA: TATTTAT before BBRF1"
     promoter        complement(113885)
                     /note="TATA: CATAAAT"
     CDS             114204..116045
                     /note="BBRF1 late reading frame, homologous to RF 54 VZV;
                     NCBI gi: 59125"
                     /codon_start=1
                     /translation="MFNMNVDESASGALGSSAIPVHPTPASVRLFEILQGKYAYVQGQ
                     TIYANLRNPGVFSRQVFTHLFKRAISHCTYDDVLHDWNKFEACIQKRWPSDDSCASRF
                     RESTFESWSTTMKLTVRDLLTTNIYRVLHSRSVLSYERYVDWICATGMVPAVKKPITQ
                     ELHSKIKSLRDRCVCRELGHERTIRSIGTELYEATKEIIESLNSTFIPQFTEVTIEYL
                     PRSDEYVAYYCGRRIRLHVLFPPAIFAGTVTFDSPVQRLYQNIFMCYRTLEHAKICQL
                     LNTAPLKAIVGHGGRDMYKDILAHLEQNSQRKDPKKELLNLLVKLSENKTISGVTDVV
                     EEFITDASNNLVDRNRLFGQPGETAAQGLKKKVSNTVVKCLTDQINEQFDQINGLEKE
                     RELYLKKIRSMESQLQASLGPGGNNPAASAPAAVAAEAASVDILTGSTASAIEKLFNS
                     PSASLGARVSGHNESILNSFVSQYIPPSREMTKDLTELWESELFNTFKLTPVVDNQGQ
                     RLYVRYSSDTISILLGPFTYLVAELSPVELVTDVYATLGIVEIIDELYRSSRLAIYIE
                     DLGRKYCPASATGGDHGIRQAPSARGDTEPDHAKSKPARDPPPGAGS"
     CDS             115843..116784
                     /note="BBRF2 late reading frame, homologous to RF 53 VZV;
                     NCBI gi: 583894"
                     /codon_start=1
                     /translation="XSWRSSTSSTGAVAWPSTSRTSVENTAPRARPGEIMASGKHHQP
                     GGTRSLTMQKVSLRVTPRLVLEVNRHNAICVATNVPEFYNARGDLNIRDLRAHVKARM
                     ISSQFCGYVLVSLLDSEDQVDHLNIFPHVFSERMILYKPNNVNLMEMCALLSMIENAK
                     SPSIGLCREVLGRLTLLHSKCNNLDSLFLYNGARTLLSTLVKYHDLEEGAATPGPWNE
```

```
                GLSLFKLHKELKRAPSEARDLMQSLFLTSGKMGCLARSPKDYCADLNKEEDANSGFTF
                NLFYQDSLLTKHFQCQTVLQTLRRKCLGSDTVSKIIP"
promoter        complement(116683)
                /note="TATA: GATAAAA"
misc_feature    complement(116696)
                /note="polyA signal: AATAAA"
CDS             complement(116781..117386)
                /note="BBLF3 early reading frame, spliced to BBLF2. BBLF3
                contains a consensus nucleotide binding site;  NCBI gi:
                583895"
                /codon_start=1
                /translation="XFLQGVKDSEDASRLDRDVMGGEATVARRHIRVKARRGPGCLLM
                AIFQGDLYVGGCREHSGPFLVWHEAFSWTLDQLAARPEADKAPPSHDHLLTLVRDLTR
                RLAPGRRRNRFWALPRAWLQRLRRAGLRLSGSHVCLLDKDGARPAPCQTATEHGLSPT
                AYFREIMAFLLDVISALHPGYTIPMEITRETDLLMTVLSLF"
misc_feature    116785
                /note="polyA signal: AATAAA"
intron          complement(117386..117515)
                /note="intron spliced out in RNA linking BBLF2 and BBLF3"
CDS             complement(117515..119080)
                /note="BBLF2 early reading frame, spliced to BBLF3;  NCBI
                gi: 583896"
                /codon_start=1
                /translation="XDVLKGPVLLRSQTMMETPAESVRARVSSVTFYNVTQTAGRWWA
                IWVVGIVPIKREDVETLIVVQACQPPLGGSLEPPVVNAPSTTELNFLRWERELRRSGG
                LIAMLADAAEKDLFDLSFRTRDRRLLSAARVEDEQGLIFQPLFPAQVVCQSCSGDDGR
                DQQPPPVDGFGSEMEGEQTCPHAQRHSESPGQLDVYIRTPRGDVFTYSTETPDDPSPV
                PFRDILRPVTYEVDLVSSDGATGRGGDARRHRVSLKILEPAGGFESWLVNSWSMAGGG
                LYAFLRSIYASCYANHRGTKPIFYLLDPELCPGGSDFQPYVPGFPFLPIHYVGRARPA
                FWHRAPHSEGLLLLDLNLGVSGTPLADALLGLDARSGQRRGSLLLQQIWPPTRKEINP
                RHVCTREGGEGGGEDETTVVGRAEATAILEADATWWLYELARCHLSARGAPVGTPDGG
                GQARDAQTWLRALHRYGTSDTRRALGGLYTAVTRVLLHAAADLGLTWAYADEFILGFV
                APTSAHPSEEPLAQ"
promoter        118981
                /note="TATA: TATAAAA BBR1 late promoter before BBRF3"
promoter        119067
                /note="TATA: TTTAAAA BBR2 late promoter ?"
promoter        119098
                /note="TATA: TATTTAA BBR3 late promoter before BBRF3"
misc_feature    119108
                /note="DONOR: AAGGTGAAT possible donor"
CDS             119137..120354
                /note="BBRF3 late reading frame;  NCBI gi: 59130"
                /codon_start=1
                /translation="MKSSKNDTFVYRTWVKTLVVYFVMFVMSAVVPITAMFPNLGYPC
                YFNALVDYGALNLTNYNLAHHLTPTLYLEPPEMFVYITLVFIADCVAFIYYACGEVAL
                IKARKKVSGLTDLSAWVSAVGSPTVLFLAILKLWSIQVFIQVLSYKHVFLSAFVYFLH
                FLASVLHACACVTRFSPVWVVKAQDNSIPQDTFLWWVVFYLKPVVTNLYLGCLALETL
                VFSLSVFLALGNSFYFMVGDMVLGAVNLFLILPIFWYILTEVWLASFLRHNFGFYCGM
                FIASIILILPLVRYEAVFVSAKLHTTVAINVAIIPILCSVAMLIRICRIFKSMRQGTD
                YVPVSETVELELESEPRPRPSRTPSPGRNRRRSSTSSSSSRSTRRQRPVSTQALVSSV
                LPMTTDSEEEIFP"
misc_feature    120260
                /note="ACCEPT: ATCTTCCTCCAGGT possible acceptor"
misc_feature    120358
                /note="polyA signal: AATAAA"
CDS             complement(120747..120974)
                /note="BBLF1 late reading frame, possibly homologous to RF
                49 VZV;  NCBI gi: 59131"
                /codon_start=1
```



```
                /translation="MGALWSLCRRRVNSIGDVDGGIINLYNDYEEFNLETTKLIAAEE
                GRACGETNEGLEYDEDSENDELLFLPNKKPN"
misc_feature    complement(120764)
                /note="polyA signal: AATAAA, 3' end of 0.6kb late, 1.6kb
                early, 3.0kb early RNAs"
CDS             complement(120929..122341)
                /note="BGLF5 early reading frame, homologous to RF 48 VZV
                and alkaline exonuclease of HSV;  NCBI gi: 59132"
                /codon_start=1
                /translation="MADVDELEDPMEEMTSYTFARFLRSPETEAFVRNLDRPPQMPAM
                RFVYLYCLCKQIQEFSGETGFCDFVSSLVQENDSKDGPSLKSIYWGLQEATDEQRTVL
                CSYVESMTRGQSENLMWDILRNGIISSSKLLSTIKNGPTKVFEPAPISTNHYFGGPVA
                FGLRCEDTVKDIVCKLICGDASANRQFGFMISPTDGIFGVSLDLCVNVESQGDFILFT
                DRSCIYEIKCRFKYLFSKSEFDPIYPSYTALYKRPCKRSFIRFINSIARPTVEYVPDG
                RLPSEGDYLLTQDEAWNLKDVRKRKLGPGHDLVADSLAANRGVESMLYVMTDPSENAG
                RIGIKDRVPVNIFINPRHNYFYQVLLQYKIVGDYVRHSGGGKPGRDCSPRVNIVTAFF
                RKRSPLDPATCTLGSDLLLDASVEIPVAVLVTPVVLPDSVIRKTLSTAAGSWKAYADN
                TFDTAPWVPSGLFADDESTP"
promoter        complement(121331)
                /note="TATA: TATTAAA BBL1 late promoter before BBLF1"
promoter        121669
                /note="TATA: CATAAAT"
promoter        121697
                /note="TATA: TATAAAG"
promoter        121772
                /note="TATA: CATAAAG"
misc_feature    122313
                /note="BAM: Bam H1 B/G"
CDS             complement(122325..123692)
                /note="BGLF4 early reading frame, homologous to RF 47 VZV;
                 NCBI gi: 583897"
                /codon_start=1
                /translation="XGWRSSVSRSLRPETSCDRPSSHLRNMDVNMAAELSPTNSSSSG
                ELSVSPEPPRETQAFLGKVTVIDYFTFQHKHLKVTNIDDMTETLYVKLPENMTRCDHL
                PITCEYLLGRGSYGAVYAHADNATVKLYDSVTELYHELMVCDMIQIGKATAEDGQDKA
                LVDYLSACTSCHALFMPQFRCSLQDYGHWHDGSIEPLVRGFQGLKDAVYFLNRHCGLF
                HSDISPSNILVDFTDTMWGMGRLVLTDYGTASLHDRNKMLDVRLKSSKGRQLYRLYCQ
                REPFSIAKDTYKPLCLLSKCYILRGAGHIPDPSACGPVGAQTALRLDLQSLGYSLLYG
                IMHLADSTHKIPYPNPDMGFDRSDPLYFLQFAAPKVVLLEVLSQMWNLNLDMGLTSCG
                ESPCVDVTAEHMSQFLQWCRSLKKRFKESYFFNCRPRFEHPHLPGLVAELLADDFFGP
                DGRRG"
misc_feature    complement(123506)
                /note="DONOR: AAGGTGACT possible donor"
CDS             complement(123941..124939)
                /note="BGLF3 reading frame;  NCBI gi: 59135"
                /codon_start=1
                /translation="MFNAVKADMPDDPMLARRYGQCLELALEACQDTPEQFKLVETPL
                KSFLLVSNILPQDNRPWHEARSSGRVAEDDYDFSSLALELLPLNPRLPEEWQFGGQGW
                SSRMEPSQPEMGMGLCFEVFDGDLMRIALAWNKDEVIGQALQILAHSQTWTSLVPEDP
                LPWMWALFYGPRSHCEERHCVYAAARGKRGPILLPTAVYTPCANIEAFLAHLTRCVYA
                LYLDVRDWKGEDIAPPFDVSRLNKMAKQLCLLPQEPFCITRVCLLCLLHKQNLNAQYK
                RPVDTYDPCLILTGEAERYMVDAVGNYREASTGTTVLYPTYDLGSIVADMVTYEDE"
promoter        complement(124117)
                /note="TATA: TATAAAA"
misc_feature    complement(124219)
                /note="polyA signal: AATAAA"
CDS             124938..125915
                /note="BGRF1 reading frame, homologous to RF 45 VZV and
                spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
                Northern blots in BGRF1 detect 2.7, 2.6, 2.1kb late and
```

```
                     1.9kb early RNAs. 2.6, 2.1kb RNAs very weak.  NCBI gi:
                     59134"
                     /codon_start=1
                     /translation="MLYASQRGRLTENLRNALQQDSTTQGCLGAETPSIMYTGAKSDR
                     WAHPLVGTIHASNLYCPMLRAYCRHYGPRPVFVASDESLPMFGASPALHTPVQVQMCL
                     LPELRDTLQRLLPPPNLEDSEALTEFKTSVSSARAILEDPNFLEMREFVTSLASFLSG
                     QYKHKPARLEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMTLEKLHIFK
                     QKASVFLIPRRHGKTWIVVAIISLILSNLSNVQIGYVAHQKHVASAVFTEIIDTLTKS
                     FDSKRVEVNKETSTITFRHSGKISSTVMCATCFNKNVRPDVSVLGNCRA"
promoter             complement(125113)
                     /note="TATA: TATAAAT before BGLF3"
misc_feature         complement(125484)
                     /note="polyA signal: AATAAA, 3' end of 1.6kb late, 1.8kb
                     late, 3.0kb late and 3.7kb early RNAs"
CDS                  complement(125863..126873)
                     /note="BGLF2 late reading frame, poor homology to RF 44
                     VZV;  NCBI gi: 59136"
                     /codon_start=1
                     /translation="MASAANSSREQLRKFLNKECLWVLSDASTPQMKVYTATTAVSAV
                     YVPQIAGPPKTYMNVTLIVLKPKKKPTYVTVYINGTLATVARPEVLFTKAVQGPHSLT
                     LMYFGVFSDAVGEAVPVEIRGNPVVTCTDLTTAHVFTTSTAVKTVEELQDITPSEIIP
                     LGRGGAWYAEGALYMFFVNMDMLMCCPNMPTFPSLTHFINLLTRCDNGECVTCYGAGA
                     HVNILRGWTEDDSPGTSGTCPCLLPCTALNNDYVPITGHRALLGLMFKPEDAPFVVGL
                     RFNPPKMHPDMSRVLQGVLANGKEVPCTAQPWTLLRFSDLYSRAMLYNCQVLKRQVLH
                     SY"
promoter             126277
                     /note="TATA: GATAAAA"
CDS                  complement(126851..128374)
                     /note="BGLF1 late reading frame;  NCBI gi: 59137"
                     /codon_start=1
                     /translation="MDVHIDNQVLSGLGTPLLVHLFVPDTVMAELCPNRVPNCEGAWC
                     QTLFSDRTGLTRVCRVFAARGMLPGRPSHRGTFTSVPVYCDEGLPELYNPFHVAALRF
                     YDEGGLVGELQIYYLSLFEGAKRALTDGHLIREASGVQESAAAMQPIPIDPGPPGGAG
                     IEHMPVAAAQVEHPKTYDLKQILLEITQEENRGEQRLGHAGSPALCLGLRLRAGAETK
                     AAAETSVSKHHPALENPSNIRGSAGGEGGGGRAGTGGTVGVGSGALSRVPVSFSKTRR
                     AIRESRALVRGIAHIFSPHALYVVTYPELSAQGRLHRMTAVTHASPATDLAEVSILGA
                     PEREFRFLISVALRISASFREKLAMQAWTAQQEIPVVIPTSYSRIYKNSDLIREAFFT
                     VQTRVSWESCWVKATISNAPKTPDACLWIDSHPLYEEGASAWGKVIDSRPPGGLVGAA
                     SQLVALGTDGHCVHLATTSDGQAFLVLPGGFVIKGQLALTPEERGYILARHGIRREQ"
promoter             complement(126929)
                     /note="TATA: TATTAAA EEL8 late promoter before BGLF2,
                     gives 1.6kb late RNA"
promoter             complement(127237)
                     /note="TATA: TATAAAA, potential promoter for 1.8kb late
                     RNA"
misc_feature         128029
                     /note="polyA signal: AATAAA"
CDS                  complement(128344..129021)
                     /note="BDLF4 early reading frame;  NCBI gi: 59138"
                     /codon_start=1
                     /translation="MSDQGRLSLPRGEGGTDEPNPRHLCSYSKLEFHLPLPESMASVF
                     ACWGCGEYHVCDGSSECTLIETHEGVVCALTGNYMGPHFQPALRPWTEIRQDTQDQRD
                     KWEPEQVQGLVKTVVNHLYHYFLNENVISGVSEALFDQEGALRPHIPALVSFVFPCCL
                     MLFRGASSEKVVDVVLSLYIHVIISIYSQKTVYGALLFKSTRNKRYDAVAKRMRELWM
                     STLTTKC"
promoter             complement(128432)
                     /note="TATA: TATTTAA before BGLF1, potential promoter for
                     3.0kb late RNA"
misc_feature         128848
                     /note="BAM: Bam H1 G/D"
```



```
promoter             complement(129054)
                     /note="TATA: TATTTGC before BDLF4, potential promoter for
                     3.7kb early RNA"
CDS                  129188..130351
                     /note="BDRF1 reading frame, homologous to RF 42 VZV and
                     spliced gene in HSV (Costa et al, 1985). Spliced from
                     BGRF1. Northern blots in BDRF1 detect 2.7, 2.6 kb late and
                     1.9kb early RNAs. Possibly also 1.8kb early RNA.  NCBI gi:
                     583898"
                     /codon_start=1
                     /translation="XLASYPVFQSIRGQTFHLLFVDEANFIKKEALPAILGFMLQKDA
                     KIIFISSVNSADQATSFLYKLKDAQERLLNVVSYVCQEHRQDFDMQDSMVSCPCFRLH
                     IPSYITMDSNIRATTNLFLDGAFSTELMGDTSSLSQGSLSRTVRDDAINQLELCRVDT
                     LNPRVAGRLASSLYVYVDPAYTNNTSASGTGIAAVTHDRADPNRVIVLGLEHFFLKDL
                     TGDAALQIATCVVALVSSIVTLHPHLEEVKVAVEGNSSQDSAVAIASIIGESCPLPCA
                     FVHTKDKTSSLQWPMYLLTNEKSKAFERLIYAVNTASLSASQVTVSNTIQLSFDPVLY
                     LISQIRAIKPIPLRDGTYTYTGKQRNLSDDVLVALVMAHFLATTQKHTFKKVH"
promoter             129374
                     /note="TATA: TATAAGC"
promoter             complement(129377)
                     /note="TATA: TATAAAG"
misc_feature         129413
                     /note="DONOR: GTGGTAAGT possible donor"
misc_feature         130347
                     /note="polyA signal: ATTAAA"
misc_feature         complement(130359)
                     /note="polyA signal: AATAAA, 3' end of 0.9kb late RNA,
                     2.3kb late RNA and 3.2kb late RNA"
CDS                  complement(130362..131066)
                     /note="BDLF3 late reading frame 9xNXT/S;  NCBI gi: 59140"
                     /codon_start=1
                     /translation="MAHARDKAGAVMAMILICETSLIWTSSGSSTASAGNVTGTTAVT
                     TPSPSASGPSTNQSTTLTTTSAPITTTAILSTNTTTVTFTGTTVTPVPTTSNASTINV
                     TTKVTAQNITATEAGTGTSTGVTSNVTTRSSSTTSATTRITNATTLAPTLSSKGTSNA
                     TKTTAELPTVPDERQPSLSYGLPLWTLVFVGLTFLMLILIFAAGLMMSAKNKPLDEAL
                     LTNAVTRDPSLYKGLV"
promoter             complement(131104)
                     /note="TATA: TATAAAA EEL4 late promoter before BDLF3,
                     gives 0.9kb late RNA"
CDS                  complement(131127..132389)
                     /note="BDLF2 late reading frame;  NCBI gi: 59141"
                     /codon_start=1
                     /translation="MVDEQVAVEHGTVSHTISREEDGVVHERRVLASGERVEVFYKAP
                     APRPREGRASTFHDFTVPAAAAVPGPEPEPEPHPPMPIHANGGGETKTNTQDQNQNQT
                     TRTRTNAKAEERTAEMDDTMASSGGQRGAPISADLLSLSSLTGRMAAMAPSWMKSEVC
                     GERMRFKEDVYDGEAETLAEPPRCFMLSFVFIYYCCYLAFLALLAFGFNPLFLPSFMP
                     VGAKVLRGKGRDFGVPLSYGCPTNPFCKVYTLIPAVVINNVTYYPNNTDSHGGHGGFE
                     AAALHVAALFESGCPNLQAVTNRNRTFNVTRASGRVERRLVQDMQRVLASAVVVMHHH
                     CHYETYYVFDGVGPEFGTIPTPCFKDVLAFRPSLVTNCTAPLKTSVKGPNWSGAAGGM
                     KRKQCRVDRLTDRSFPAYLEEVMYVMVQ"
promoter             132266
                     /note="TATA: TATAAAA"
CDS                  complement(132400..133305)
                     /note="BDLF1 late reading frame, poor homology to RF 41
                     VZV;  NCBI gi: 59142"
                     /codon_start=1
                     /translation="MDLKVVVSLSSRLYTDEIAKMQQRIGCILPLASTHGTQNVQGLG
                     LGQVYSLETVPDYVSMYNYLSDCTLAVLDEVSVDSLILTKIVPGQTYAIKNKYQPFFQ
                     WHGTGSLSVMPPVFGREHATVKLESNDVDIVFPMVLPTPIAEEVLQKILLFNVYSRVV
                     MQAPGNADMLDVHMHLGSVSYLGHHYELALPEVPGPLGLALLDNLSLYFCIMVTLLPR
```

ASMRLVRGLIRHEHHDLLNLFQEMVPDEIARIDLDDLSVADDLSRMRVMMTYLQSLAS
LFNLGPRLATAAYSQETLTATCWLR"
promoter complement(132476)
/note="TATA: TATTTAA before BDLF2, likely promoter for
2.3kb late RNA"
misc_feature complement(133312)
/note="polyA signal: AATAAA, 3' end of 4.5kb late RNA"
CDS complement(133321..137466)
/note="BcLF1 late reading frame, homologous to RF 40 VZV
and major capsid protein of HSV; NCBI gi: 59143"
/codon_start=1
/translation="MASNEGVENRPFPYLTVDADLLSNLRQSAAEGLFHSFDLLVGKD
AREAGIKFEVLLGVYTNAIQYVRFLETALAVSCVNTEFKDLSRMTDGKIQFRISVPTI
AHGDGRRPSKQRTFIVVKNCHKHHISTEMELSMLDLEILHSIPETPVEYAEYVGAVKT
VASALQFGVDALERGLINTVLSVKLRHAPPMFILQTLADPTFTERGFSKTVKSDLIAM
FKRHLLEHSFFLDRAENMGSGFSQYVRSRLSEMVAAVSGESVLKGVSTYTTAKGGEPV
GGVFIVTDNVLRQLLTFLGEEADNQIMGPSSYASFVVRGENLVTAVSYGRVMRTFEHF
MARIVDSPEKAGSTKSDLPAVAAGVEDQPRVPISAAVIKLGNHAVAVESLQKMYNDTQ
SPYPLNRRMQYSYYFPVGLFMPNPKYTTSAAIKMLDNPTQQLPVEAWIVNKNNLLLAF
NLQNALKVLCHPRLHTPAHTLNSLNAAPAPRDRRETYSLQHRRPNHMNVLVIVDEFYD
NKYAAPVTDIALKCGLPTEDFLHPSNYDLLRLELHPLYDIYIGRDAGERARHRAVHRL
MVGNLPTPLAPAAFQEARGQQFETATSLAHVVDQAVIETVQDTAYDTAYPAFFYVVEA
MIHGFEEKFVMNVPLVSLCINTYWERSGRLAFVNSFSMIKFICRHLGNNAISKEAYSM
YRKIYGELIALEQALMRLAGSDVVGDESVGQYVCALLDPNLLPPVAYTDIFTHLLTVS
DRAPQIIIGNEVYADTLAAPQFIERVGNMDEMAAQFVALYGYRVNGDHDHDFRLHLGP
YVDEGHADVLEKIFYYVFLPTCTNAHMCGLGVDFQHVAQTLAYNGPAFSHHFTRDEDI
LDNLENGTLRDLLEISDLRPTVGMIRDLSASFMTCPTFTRAVRVSVDNDVTQQLAPNP
ADKRTEQTVLVNGLVAFAFSERTRAVTQCLFHAIPFHMFYGDPRVAATMHQDVATFVM
RNPQQRAVEAFNRPEQLFAEYREWHRSPMGKYAAECLPSLVSISGMTAMHIKMSPMAY
IAQAKLKIHPGVAMTVVRTDEILSENILFSSRASTSMFIGTPNVSRREARVDAVTFEV
HHEMASIDTGLSYSSTMTPARVAAITTDMGIHTQDFFSVFPAEAFGNQQVNDYIKAKV
GAQRNGTLLRDPRTYLAGMTNVNGAPGLCHGQQATCEIIVTPVTADVAYFQKSNSPRG
RAACVVSCENYNQEVAEGLIYDHSRPDAAYEYRSTVNPWASQLGSLGDIMYNSSYRQT
AVPGLYSPCRAFFNKEELLRNNRGLYNMVNEYSQRLGGHPATSNTEVQFVVIAGTDVF
LEQPCSFLQEAFPALSASSRALIDEFMSVKQTHAPIHYGHYIIEEVAPVRRILKFGNK
VVF"
misc_feature complement(133332)
/note="DONOR: AAGGTGGTT possible donor"
promoter complement(133352)
/note="TATA: TATTAAA before BDLF1"
promoter complement(133386)
/note="TATA: TATATAA"
misc_feature 135178
/note="polyA signal: AATAAA"
promoter 135394
/note="TATA: TATAAGT"
misc_feature 136624
/note="polyA signal: AATAAA"
misc_feature 136868
/note="BAM: Bam H1 D/c"
promoter complement(137710)
/note="TATA: TATTAAA EHL1 promoter before BcLF1, gives
4.5kb late RNA"
promoter 137857
/note="TATA: CATAAAC"
CDS 137862..139718
/note="BcRF1 reading frame; NCBI gi: 583899"
/codon_start=1
/translation="XRAQCFALSSATCLITNLSAASSSSAAWSRVGFLGTCKRNSAKM
LAHLNQVTRIPPCPPFSGREARLKFHFFSWSTFMLSWPNNATLREIRTRAATNLTHHP

HLVDTLYHASPQTPFLTRSGALYRFVTCCNCTLPNISIQQCKAGDRPGDLEIILQSNG
GGRPASFQFPSSPTGSLLRCIVAASLLPEVSVGHQELSPLRSRSQGGQTDVRSGPDPA
RRLVALLRREDGAPKDPPLGPFGHPRGPGPAKSEDEESERRDAPPPPLDSSFQASRLV
PVGPGFRLLVFNTNRVINTKLVCSEPLVKMRVCNVPRLINNFVARKYVVKETAFTVSL
FFTDGVGANLAINVNISGTYLSFLLAMTSLRCFLPVEAIYPAAVSNWNSTLDLHGLEN
QSLVRENRSGVFWTTNFPSVVSCRDGLNVSWFKAATATISRVHGQTLEQHLIREITPI
VTHREAKISRIKNRLFTLLELRNRSQIQVLHKRFLEGLLDCASLLRLDPSCINRIASE
GLFDFSKRSIAHSKNRHECALLGHRHSANVTKLVVNERKTRLDILGRNANFLTRCKHQ
VNLRQSPIFLTLLRHIRRRLGLGRASVKREITLLLAHLRKKTAPIHCRDAQV"
misc_feature 138019
/note="BAM: Bam H1 c/b"
misc_feature 139352
/note="BAM: Bam H1 b/T"
CDS 139642..140919
/note="BTRF1 reading frame. Northern blots detect 0.95
late and 3.8kb early RNA; NCBI gi: 583900"
/codon_start=1
/translation="XERLPFSWPTCAKRQPPSTAVMLKCKQPGARFIHGAVHLPSGQI
VFHTIHSPTLASALGLPGENVPIPALFRASGLNVRESLPMTNMRAPIISLARLILAPN
PYILEGQLTVGMTQDNGIPVLFARPVIEVKSGPESNIKASSQLMIAEDSCLNQIAPFS
ASEHPAFSMVESVKRVRVDEGANTRRTIRDILEIPVTVLSSLQLSPTKSILKKAPEPP
PPEPQATFDATPYARIFYDIGRQVPKLGNAPAAQVSNVLIANRSHNSLRLVPNPDLLP
LQHLYLKHVVLKSLNLENIVQDFEAIFTSPSDTISEAETKAFEKLVEQAKNTVENIVF
CLNSICSTSTLPDVVPDVNNPNISLALEKYFLMFPPSGTIMRNVRFATPIVRLLCQGA
ELGTMAQFLGKYIKVKKETGMYTLVKLYYLLRI"
misc_feature complement(140902)
/note="polyA signal: AATAAA, 3' end of 2.5kb late RNA"
CDS complement(140916..143036)
/note="BXLF2 late reading frame, encodes gp85; homologous
to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV);
NCBI gi: 59146"
/codon_start=1
/translation="MQLLCVFCLVLLWEVGAASLSEVKLHLDIEGHASHYTIPWTELM
AKVPGLSPEALWREANVTEDLASMLNRYKLIYKTSGTLGIALAEPVDIPAVSEGSMQV
DASKVHPGVISGLNSPACMLSAPLEKQLFYYIGTMLPNTRPHSYVFYQLRCHLSYVAL
SINGDKFQYTGAMTSKFLMGTYKRVTEKGDEHVLSLVFGKTKDLPDLRGPFSYPSLTS
AQSGDYSLVIVTTFVHYANFHNYFVPNLKDMFSRAVTMTAASYARYVLQKLVLLEMKG
GCREPELDTETLTTMFEVSVAFFKVGHAVGETGNGCVDLRWLAKSFFELTVLKDIIGI
CYGATVKGMQSYGLERLAAMLMATVKMEELGHLTTEKQEYALRLATVGYPKAGVYSGL
IGGATSVLLSAYNRHPLFQPLHTVMRETLFIGSHVVLRELRLNVTTQGPNLALYQLLS
TALCSALEIGEVLRGLALGTESGLFSPCYLSLRFDLTRDKLLSMAPQEATLDQAAVSN
AVDGFLGRLSLEREDRDAWHLPAYKCVDRLDKVLMIIPLINVTFIISSDREVRGSALY
EASTTYLSSSLFLSPVIMNKCSQGAVAGEPRQIPKIQNFTRTQKSCIFCGFALLSYDE
KEGLETTTYITSQEVQNSILSSNYFDFDNLHVHYLLLTTNGTVMEIAGLYEERAHVVL
AIILYFIAFALGIFLVHKIVMFFL"
misc_feature 140970
/note="polyA signal: AATAAA"
misc_feature complement(141286)
/note="polyA signal: AATAAA"
promoter 142589
/note="TATA: GATAAAA"
misc_feature 142740
/note="BAM: Bam H1 T/X"
CDS complement(143038..144861)
/note="BXLF1 early reading frame, thymidine kinase
(Littler et al, 1986). Weak homology to RF 36 VZV and HSV
thymidine kinase. 4.0kb early RNA presumably encodes the
TK. Also a 2.2kb late RNA here. NCBI gi: 59148"
/codon_start=1
/translation="MAGFPGKEAGPPGGWRKCQEDESPENERHENFYAEIDDFAPSVL

```
                TPTGSDSGAGEEDDDGLYQVPTHWPPLMAPTGLSGERVPCRTQAAVTSNTGNSPGSRH
                TSCPFTLPRGAQPPAPAHQKPTAPTPKPRSRECGPSKTPDPFSWFRKTSCTEGGADST
                SRSFMYQKGFEEGLAGLGLDDKSDCESEDESNFRRPSSHSALKQKNGGKGKPSGLFEH
                LAAHGREFSKLSKHAAQLKRLSGSVMNVLNLDDAQDTRQAKAQRKESMRVPIVTHLTN
                HVPVIKPACSLFLEGAPGVGKTTMLNHLKAVFGDLTIVVPEPMRYWTHVYENAIKAMH
                KNVTRARHGREDTSAEVLACQMKFTTPFRVLASRKRSLLVTESGARSVAPLDCWILHD
                RHLLSASVVFPLMLLRSQLLSYSDFIQVLATFTADPGDTIVWMKLNVEENMRRLKKRG
                RKHESGLDAGYLKSVNDAYHAVYCAWLLTQYFAPEDIVKVCAGLTTITTVCHQSHTPI
                IRSGVAEKLYKNSIFSVLKEVIQPFRADAVLLEVCLAFTRTLAYLQFVLVDLSEFQDD
                LPGCWTEIYMQALKNPAIRSQFFDWAGLSKVISDFERGNRD"
promoter        complement(143310)
                /note="TATA: TATAAGA ECL2 late promoter before BXLF2,
                gives 2.5kb late RNA"
misc_feature    143608
                /note="polyA signal: AATAAA"
misc_feature    144791
                /note="ACCEPT: TCTTTCGTTTTCAGG poss. acceptor before
                BXRF1"
CDS             144860..145606
                /note="BXRF1 late reading frame, homologous to RF 35 VZV.
                Basic (core?) protein.  NCBI gi: 59147"
                /codon_start=1
                /translation="MDPTRGLCALSTHDLAKFHSLPPARKAAGKRAHLRCYSKLLSLK
                SWEQLASFLSLPPGPTFTDFRLFFEVTLGRRIADCVVVALQPYPRCYIVEFKTAMSNT
                ANPQSVTRKAQRLEGTAQLCDCANFLRTSCPPVLGSQGLEVLAALVFKNQRSLRTLQV
                EFPALGQKTLPTSTTGLLNLLSRWQDGALRARLDRPRPTAQGHRPRTHVGPKPSQLTA
                RVPRSARAGRAGGRKGQVGAVGQVCPGAQK"
misc_feature    144862
                /note="BAM: Bam H1 X/V"
misc_feature    144945
                /note="DONOR: CAGGTAAGC possible donor at 3' BXRF1"
promoter        complement(145135)
                /note="TATA: TATAACA before BXLF1"
promoter        145302
                /note="TATA: TATTTAA before BVRF1, potential promoter for
                1.9kb early RNA"
CDS             145416..147128
                /note="BVRF1 early reading frame, homologous to RF 34 VZV;
                 NCBI gi: 59149"
                /codon_start=1
                /translation="MALSGHVLIDPARLPRDTGPELMWAPSLRNSLRVSPEALELAER
                EAERARSERWDRCAQVLKNRLLRVELDGIMRDHLARAEEIRQDLDAVVAFSDGLESMQ
                VRSPSTGGRSAPAPPSPSPAQPFTRLTGNAQYAVSISPTDPPLMVAGSLAQTLLGNLY
                GNINQWVPSFGPWYRTMSANAMQRRVFPKQLRGNLNFTNSVSLKLMTEVVAVLEGTTQ
                DFFSDVRHLPDLQAALILSVAYLLLQGGSSHQQRPLPASREELLELGPESLEKIIADL
                KAKSPGGNFMILTSGNKEARQSIAPLNRQAAYPPGTFADNKIYNLFVGAGLLPTTAAL
                NVPGAAGRDRDLVYRIANQIFGEDVPPFSSHQWNLRVGLAALEALMLVYTLCETANLA
                EAATRRLHLSSLLPQAMQRRKPAMASAGMPGAYPVQTLFRHGELFRFIWAHYVRPTVA
                ADPQASISSLFPGLVLLALELKLMDGQAPSHYAINLTGQKFDTLFEIINQKLLFHDPA
                AMLAARTQLRLAFEDGVGVALGRPSPMLAAREILERQFSASDDYDRLYFLTLGYLASP
                VAPS"
misc_feature    complement(146926)
                /note="polyA signal: AATAAA"
misc_feature    147167
                /note="DONOR: AAGGTAAAT possible donor"
misc_feature    147170
                /note="polyA signal: AATAAA, 3' end of 2.4kb late and
                1.9kb early RNAs"
promoter        147721
                /note="TATA: TATTTAT before BVRF2, potential promoter for
```



```
                2.1kb early RNA"
CDS             147927..149744
                /note="BVRF2 early reading frame, N-terminus homologous to
                RF 33 VZV;  NCBI gi: 59150"
                /codon_start=1
                /translation="MVQAPSVYVCGFVERPDAPPKDACLHLDPLTVKSQLPLKKPLPL
                TVEHLPDAPVGSVFGLYQSRAGLFSAASITSGDFLSLLDSIYHDCDIAQSQRLPLPRE
                PKVEALHAWLPSLSLASLHPDIPQTTADGGKLSFFDHVSICALGRRRGTTAVYGTDLA
                WVLKHFSDLEPSIAAQIENDANAAKRESGCPEDHPLPLTKLIAKAIDAGFLRNRVETL
                RQDRGVANIPAESYLKASDAPDLQKPDKALQSPPPASTDPATMLSGNAGEGATACGGS
                AAAGQDLISVPRNTFMTLLQTNLDNKPPRQTPLPYAAPLPPFSHQAIATAPSYGPGAG
                AVAPAGGYFTSPGGYYAGPAGGDPGAFLAMDAHTYHPHPHPPPAYFGLPGLFGPPPPV
                PPYYGSHLRADYVPAPSRSNKRKRDPEEDEEGGGLFPGEDATLYRKDIAGLSKSVNEL
                QHTLQALRRETLSYGHTGVGYCPQQGPCYTHSGPYGFQPHQSYEVPRYVPHPPPPPTS
                HQAAQAQPPPPGTQAPEAHCVAESTIPEAGAAGNSGPREDTNPQQPTTEGHHRGKKLV
                QASASGVAQSKEPTTPKAKSVSAHLKSIFCEELLNKRVA"
misc_feature    148007
                /note="BAM: Bam H1 V/d"
promoter        148620
                /note="TATA: TATTTAA ECR1 late promoter before BdRF1,
                gives 1.2kb late RNA"
CDS             148707..149744
                /note="BdRF1 reading frame; this is the C terminus of
                BVRF2;  NCBI gi: 59151"
                /codon_start=1
                /translation="MLSGNAGEGATACGGSAAAGQDLISVPRNTFMTLLQTNLDNKPP
                RQTPLPYAAPLPPFSHQAIATAPSYGPGAGAVAPAGGYFTSPGGYYAGPAGGDPGAFL
                AMDAHTYHPHPHPPPAYFGLPGLFGPPPPVPPYYGSHLRADYVPAPSRSNKRKRDPEE
                DEEGGGLFPGEDATLYRKDIAGLSKSVNELQHTLQALRRETLSYGHTGVGYCPQQGPC
                YTHSGPYGFQPHQSYEVPRYVPHPPPPPTSHQAAQAQPPPPGTQAPEAHCVAESTIPE
                AGAAGNSGPREDTNPQQPTTEGHHRGKKLVQASASGVAQSKEPTTPKAKSVSAHLKSI
                FCEELLNKRVA"
misc_feature    149115
                /note="BAM: Bam H1 d/I"
misc_feature    149727
                /note="polyA signal: AATAAA, 3' end of 2.1kb early and
                1.2kb late RNAs"
misc_feature    complement(149758)
                /note="polyA signal: AATAAA, 3' end of 1.0kb late, 1.5kb
                late and 1.8kb late RNAs"
CDS             complement(149779..150525)
                /note="BILF2 late reading frame 11xNXT/S;  NCBI gi: 59152"
                /codon_start=1
                /translation="MTHLVLLLCCCVGSVCAFFSDLVKFENVTAHAGARVNLTCSVPS
                NESVSRIELGRGYTPGDGQLPLAVATSNNGTHITNGGYNYSLTLEWVNDSNTSVSLII
                PNVTLAHAGYYTCNVTLRNCSVASGVHCNYSAGEEDDQYHANRTLTQRMHLTVIPATT
                IAPTTLVSHTTSTSHRPHRRPVSKRPTHKPVTLGPFPIDPWRPKTTWVHWALLLITCA
                VVAPVLLIIIISCLGWLAGWGRRRKGWIPL"
promoter        complement(150571)
                /note="TATA: TATTTAG before BILF2. Potential promoter for
                1.0kb late RNA."
repeat_region   151236..151618
                /note="repetitive sequence 3X"
misc_feature    151767
                /note="polyA signal: AATAAA"
promoter        complement(151780)
                /note="TATA: CATAAAA"
misc_feature    152012..152013
                /note="DEL: B95-8 deletion with respect to Raji"
CDS             complement(152161..153099)
```

```
                     /note="BILF1 reading frame, membrane protein, 3xNXS/T;
                     NCBI gi: 59153"
                     /codon_start=1
                     /translation="MLSTMAPGSTVGTLVANMTSVNATEDACTKSYSAFLSGMTSLLL
                     VLLILLTLAGILFIIFVRKLVHRMDVWLIALLIELLLWVLGKMIQEFSSTGLCLLTQN
                     MMFLGLMCSVWTHLGMALEKTLALFSRTPKRTSHRNVCLYLMGVFCLVLLLIIILLIT
                     MGPDANLNRGPNMCREGPTKGMHTAVQGLKAGCYLLAAVLIVLLTVIIIWKLLRTKFG
                     RKPRLICNVTFTGLICAFSWFMLSLPLLFLGEAGSLGFDCTESLVARYYPGPAACLAL
                     LLIILYAWSFSHFMDSLKNQVTVTARYFRRVPSQST"
promoter             152230
                     /note="TATA: CATAAAA"
misc_feature         153259
                     /note="polyA signal: AATAAA"
misc_feature         153637
                     /note="HPN: 22bp 2-fold symmetric"
misc_feature         complement(153690)
                     /note="DONOR: AAAGTGAGG possible donor"
CDS                  complement(153699..156746)
                     /note="BALF5 DNA polymerase (early), homologous to many
                     DNA polymerases, CMV HFLF2 and RF 28 VZV. 4.5kb early RNA
                     apparently encodes BALF5, RNA ends unknown.  NCBI gi:
                     59154"
                     /codon_start=1
                     /translation="MSGGLFYNPFLRPNKGLLKKPDKEYLRLIPKCFQTPGAAGVVDV
                     RGPQPPLCFYQDSLTVVGGDEDGKGMWWRQRAQEGTARPEADTHGSPLDFHVYDILET
                     VYTHEKCAVIPSDKQGYVVPCGIVIKLLGRRKADGASVCVNVFGQQAYFYASAPQGLD
                     VEFAVLSALKASTFDRRTPCRVSVEKVTRRSIMGYGNHAGDYHKITLSHPNSVCHVAT
                     WLQDKHGCRIFEANVDATRRFVLDNDFVTFGWYSCRRAIPRLQHRDSYAELEYDCEVG
                     DLSVRREDSSWPSYQALAFDIECLGEEGFPTATNEADLILQISCVLWSTGEEAGRYRR
                     ILLTLGTCEDIEGVEVYEFPSELDMLYAFFQLIRDLSVEIVTGYNVANFDWPYILDRA
                     RHIYSINPASLGKIRAGGVCEVRRPHDAGKGFLRANTKVRITGLIPIDMYAVCRDKLS
                     LSDYKLDTVARHLLGAKKEDVHYKEIPRLFAAGPEGRRRLGMYCVQDSALVMDLLNHF
                     VIHVEVAEIAKIAHIPCRRVLDDGQQIRVFSCLLAAAQKENFILPMPSASDRDGYQGA
                     TVIQPLSGFYNSPVLVVDFASLYPSIIQAHNLCYSTMITPGEEHRLAGLRPGEDYESF
                     RLTGGVYHFVKKHVHESFLASLLTSWLAKRKAIKKLLAACEDPRQRTILDKQQLAIKC
                     TCNAVYGFTGVANGLFPCLSIAETVTLQGRTMLERAKAFVEALSPANLQALAPSPDAW
                     APLNPEGQLRVIYGDTDSLFIECRGFSESETLRFADALAAHTTRSLFVAPISLEAEKT
                     FSCLMLITKKRYVGVLTDGKTLMKGVELVRKTACKFVQTRCRRVLDLVLADARVKEAA
                     SLLSHRPFQESFTQGLPVGFLPVIDILNQAYTDLREGRVPMGELCFSTELSRKLSAYK
                     STQMPHLAVYQKFVERNEELPQIHDRIQYVFVEPKGGVKGARKTEMAEDPAYAERHGV
                     PVAVDHYFDKLLQGAANILQCLFDNNSGAALSVLQNFTARPPF"
misc_feature         154747
                     /note="BAM: Bam H1 I/A"
misc_feature         complement(156707)
                     /note="polyA signal: AATAAA; 3' end of 2.5kb late (gB) RNA
                     and 1.8kb late RNA"
CDS                  complement(156749..159322)
                     /note="BALF4 late reading frame 9xNXT/S homologous to HSV1
                     glycoprotein B (Pellet et al, 1985), CMV HFLF1 and RF 31
                     VZV (gpII);  NCBI gi: 59155"
                     /codon_start=1
                     /translation="MTRRRVLSVVVLLAALACRLGAQTPEQPAPPATTVQPTATRQQT
                     SFPFRVCELSSHGDLFRFSSDIQCPSFGTRENHTEGLLMVFKDNIIPYSFKVRSYTKI
                     VTNILIYNGWYADSVTNRHEEKFSVDSYETDQMDTIYQCYNAVKMTKDGLTRVYVDRD
                     GVNITVNLKPTGGLANGVRRYASQTELYDAPGWLIWTYRTRTTVNCLITDMMAKSNSP
                     FDFFVTTTGQTVEMSPFYDGKNKETFHERADSFHVRTNYKIVDYDNRGTNPQGERRAF
                     LDKGTYTLSWKLENRTAYCPLQHWQTFDSTIATETGKSIHFVTDEGTSSFVTNTTVGI
                     ELPDAFKCIEEQVNKTMHEKYEAVQDRYTKGQEAITYFITSGGLLLAWLPLTPRSLAT
                     VKNLTELTTPTSSPPSSPSPPAPSAARGSTPAAVLRRRRRDAGNATTPVPPTAPGKSL
                     GTLNNPATVQIQFAYDSLRRQINRMLGDLARAWCLEQKRQNMVLRELTKINPTTVMSS
```



```
                     IYGKAVAAKRLGDVISVSQCVPVNQATVTLRKSMRVPGSETMCYSRPLVSFSFINDTK
                     TYEGQLGTDNEIFLTKKMTEVCQATSQYYFQSGNEIHVYNDYHHFKTIELDGIATLQT
                     FISLNTSLIENIDFASLELYSRDEQRASNVFDLEGIFREYNFQAQNIAGLRKDLDNAV
                     SNGRNQFVDGLGELMDSLGSVGQSITNLVSTVGGLFSSLVSGFISFFKNPFGGMLILV
                     LVAGVVILVISLTRRTRQMSQQPVQMLYPGIDELAQQHASGEGPGINPISKTELQAIM
                     LALHEQNQEQKRAAQRAAGPSVASRALQAARDRFPGLRRRRYHDPETAAALLGEAETE
                     F"
promoter             158204
                     /note="TATA: TATAAAA"
CDS                  complement(159309..161678)
                     /note="BALF3 reading frame;  NCBI gi: 583901"
                     /codon_start=1
                     /translation="XRRGVLIGPLLRPGGQRPRNPGDHCLQRDRVDGGGRSGLSGGSV
                     GWARGSPGQLYQRGRAGQHRVAGLGRAAAGWRRAGLRGQTQAPSGHRSPRTRGLETPG
                     AAMSGLLAAAYSQVYALAVELSVCTRLDPRSLDVAAVVRNAGLLAELEAILLPRLRRQ
                     NDRACSALSLELVHLLENSREASAALLAPGRKGTRVPPLRTPSVAYSVEFYGGHKVDV
                     SLCLINDIEILMKRINSVFYCMSHTMGLESLERALDLLGRFRGVSPIPDPRLYITSVP
                     CWRCVGELMVLPNHGNPSTAEGTHVSCNHLAVPVNPEPVSGLFENEVRQAGLGHLLEA
                     EEKARPGGPEEGAVPGPGRPEAEGATRALDTYNVFSTVPPEVAELSELLYWNSGGHAI
                     GATGQGEGGGHSRLSALFARERRLALVRGACEEEALAGARLTHLFDAVAPGATERLFCG
                     GVYSSSGDAVEALKADCAAAFTAHPQYRAILQKRNELYTRLNRAMQRLGRGEEEASRE
                     SPEVPRPAGAREPGPSGALSDALKRKEQYLRQVATEGLAKLQSCLAQQSETLTETLCL
                     RVWGDVVYWELARMRNHFLYRRAFVSGPWEDRRAGEGAAFENSKYIKTHLFTQTLSSE
                     HLHALTHSLYTFITGPLAEESGLFPPPSNVALARCCDAAGTLPHQKAFLTSLIWPGIE
                     PSDWIETSFNSFYSVPGGSLASSQQILCRALREAVLTVSLYNKTWGRSLILRRADAVS
                     PGQALPPDGLYLTYDSDRPLILLYKGRGWVFKDLYALLYLHLQMRDDSA"
promoter             complement(159370)
                     /note="TATA: TATTTAA ECL1 late promoter before BALF4,
                     gives 2.5kb late RNA"
misc_feature         160966
                     /note="polyA signal: AATAAA"
misc_feature         160990
                     /note="polyA signal: ATTAAA"
misc_feature         complement(161013)
                     /note="polyA signal: AATAAA, presumed end of 3.9kb early
                     RNA"
CDS                  complement(161384..164770)
                     /note="BALF2 early reading frame, homologous to RF 29 VZV
                     and major DNA binding protein HSV. 3.9kb early RNA;  NCBI
                     gi: 59157"
                     /codon_start=1
                     /translation="MQGAQTSEDNLGSQSQPGPCGYIYFYPLATYPLREVATLGTGYA
                     GHRCLTVPLLCGITVEPGFSINVKALHRRPDPNCGLLRATSYHRDIYVFHNAHMVPPI
                     FEGPGLEALCGETREVFGYDAYSALPRESSKPGDFFPEGLDPSAYLGAVAITEAFKER
                     LYSGNLVAIPSLKQEVAVGQSASVRVPLYDKEVFPEGVPQLRQFYNSDLSRCMHEALY
                     TGLAQALRVRRVGKLVELLEKQSLQDQAKVAKVAPLKEFPASTISHPDSGAILMIVDSA
                     ACELAVSYAPAMLEASHETPASLNYDSWPLFADCEGPEARVAALHRYNASLAPHVSTQ
                     IFATNSVLYVSGVSKSTGQGKESLFNSFYMTHGLGTLQEGTWDPCRRPCFSGWGGPDV
                     TGTNGPGNYAVEHLVYAASFSPNLLARYAYYLQFCQGQKSSLTPVPETGSYVAGAAAS
                     PMCSLCEGRAPAVCLNTLFFRLRDRFPPVMSTQRRDPYVISGASGSYNETDFLGNFLN
                     FIDKEDDGQRPDDEPRYTYWQLNQNLLERLSRLGIDAEGKLEKEPHGPRDFVKMFKDV
                     DAAVDAEVVQFMNSMAKNNITYKDLVKSCYHVMQYSCNPFAQPACPIFTQLFYRSLLT
                     ILQDISLPICMCYENDNPGLGQSPPEWLKGHYQTLCTNFRSLAIDKGVLTAKEAKVVH
                     GEPTCDLPDLDAALQGRVYGRRLPVRMSKVLMLCPRNIKIKNRVVFTGENAALQNSFI
                     KSTTRRENYIINGPYMKFLNTYHKTLFPDTKLSSLYLWHNFSRRRSVPVPSGASAEEY
                     SDLALFVDGGSRAHEESNVIDVVPGNLVTYAKQRLNNAILKACGQTQFYISLIQGLVP
                     RTQSVPARDYPHVLGTRAVESAAAYAEATSSLTATTVVCAATDCLSQVCKARPVVTLP
                     VTINKYTGVNGNNQIFQAGNLGYFMGRGVDRNLLQAPGAGLRKQAGGSSMRKKFVFAT
                     PTLGLTVKRRTQAATTYEIENIRAGLEAIISQKQEEDCVFDVVCNLVDAMGEACASLT
                     RDDAEYLLGRFSVLADSVLETLATIASSGIEWTAEAARDFLEGVWGGPGAAQDNFISV
```

```
                AEPVSTASQASAGLLLGGGGQGSGGRRKRRLATVLPGLEV"
misc_feature    163978..166635
                /note="DEL: deletion in Raji"
promoter        complement(164814)
                /note="TATA: CATTTAA before BALF2, presumed promoter for
                3.9kb early RNA"
misc_feature    complement(164851)
                /note="polyA signal: AATAAA"
CDS             complement(164855..165517)
                /note="BALF1 early reading frame, 0.7kb early RNA;  NCBI
                gi: 59159"
                /codon_start=1
                /translation="MNLAIALDSPHPGLASYTILPRPFYHISLKPVSWPDETMRPAKS
                TDSVFVRTPVEAWVAPSPPDDKVAESSYLMFRAMYAVFTRDEKDLPLPALVLCRLIKA
                SLRKDRKLYAELACRTADIGGKDTHVRLIISVLRAVYNDHYDYWSRLRVVLCYTVVFA
                VRNYLDDHKSAAFVLGAIAHYLALYRRLWFARLGGMPRSLRRQFPVTWALASLTDFLK
                SL"
promoter        165442
                /note="TATA: GATAAAA"
promoter        165466
                /note="TATA: TATAAGA EDR1 early promoter before BARF1,
                gives 0.8kb early RNA"
CDS             165504..166169
                /note="BARF1 reading frame;  NCBI gi: 59158"
                /codon_start=1
                /translation="MARFIAQLLLLASCVAAGQAVTAFLGERVTLTSYWRRVSLGPEI
                EVSWFKLGPGEEQVLIGRMHHDVIFIEWPFRGFFDIHRSANTFFLVVTAANISHDGNY
                LCRMKLGETEVTKQEHLSVVKPLTLSVHSERSQFPDFSVLTVTCTVNAFPHPHVQWLM
                PEGVEPAPTAANGGVMKEKDGSLSVAVDLSLPKPWHLPVTCVGKNDKEEAHGVYVSGY
                LSQ"
promoter        complement(165713)
                /note="TATA: TATAAAG before BALF1"
misc_feature    166165
                /note="polyA signal: AATAAA 3' end of 0.8kb early RNA.
                Also 1kb late RNA in this region."
promoter        166469..166475
                /note="TATA: TTATTTT"
mRNA            166498..166916
                /note="Exon 1 of terminal protein RNA"
misc_feature    166561..166563
                /note="Likely initiator met of terminal protein"
misc_feature    166614
                /note="BAM: Bam H1 A/Nhet"
misc_feature    complement(166946)
                /note="polyA signal: AATAAA These 2 polyA sites are 3' end
                of 0.8kb early, 2.5kb late and 2.5kb latent RNAs"
misc_feature    complement(166950)
                /note="polyA signal: AATAAA"
CDS             complement(166998..167303)
                /note="BNLF2b reading frame;  NCBI gi: 583902"
                /codon_start=1
                /translation="XATMRPGRPLAGFYATLRRSFRRMSKRSKNKAKKERVPVEDRPP
                TPMPTSQRLIRRNALGGGVRPDAEDCIQRFHPLEPALGVSTKNFDLLSLRCELGWCG"
CDS             complement(167304..167486)
                /note="BNLF2a reading frame;  NCBI gi: 59161"
                /codon_start=1
                /translation="MVHVLERALLEQQSSACGLPGSSTETRPSHPCPEDPDVSRLRLL
                LVVLCVLFGLLCLLLI"
promoter        167320
                /note="TATA: CATAAAA"
```



```
promoter        complement(167525)
                /note="TATA: TATAAAA EDL2 early promoter before BNLF2a,b.
                Gives 0.8kb RNA"
mRNA            complement(168163..168965)
                /note="BNLF1 coding part of exon c of latent membrane
                protein"
repeat_region   168399..168574
                /note="5 x 33bp repeats"
intron          complement(168966..169041)
                /note="intervening sequence in BNLF1"
misc_feature    complement(168966)
                /note="ACCEPT: TTTTTTCAACGCAGT possible acceptor"
misc_feature    complement(169041)
                /note="DONOR: TTGGTAAGA donor"
mRNA            complement(169042..169128)
                /note="BNLF1 exon b of latent membrane protein mRNA"
intron          complement(169129..169206)
                /note="intervening sequence in BNLF1"
misc_feature    complement(169129)
                /note="ACCEPT: TCCTCTTTCCCCAGT acceptor"
promoter        complement(169201)
                /note="TATA: TATTACA EDL1A late promoter, gives 2.5kb late
                RNA"
misc_feature    complement(169206)
                /note="DONOR: TGAGTAAGT donor"
mRNA            complement(169207..169474)
                /note="BNLF1 exon a of latent membrane protein mRNA"
promoter        complement(169546)
                /note="TATA: TACATAAGC EDL1 promoter before BNLF1 gives
                2.5kb latent RNA (LMP)"
repeat_region   170094..170631
                /note="terminal repeat 1 538bp"
repeat_region   170632..171154
                /note="terminal repeat 2 523bp"
repeat_region   171155..171692
                /note="terminal repeat 3 538bp"
repeat_region   171693..172231
                /note="terminal repeat 4 538bp"
BASE COUNT    34054 a  52511 c  50755 g  34961 t
ORIGIN

Initial Score    =    170  Optimized Score  =    454  Significance = 10.52
Residue Identity =    50%  Matches          =    529  Mismatches   =   418
Gaps             =    105  Conservative Substitutions               =     0

             X       10          20          30          40          50
             CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
             ||| |||  ||    ||| ||  | |    ||    |  ||   ||   || |  |  |
    CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
107030    107040    107050    107060    107070    107080    107090    107100

    60        70          80          90          100         110         120
    CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
     |    | |||||  | | ||  ||| ||||  ||  |    ||| | |   | || ||      |  |||||
    GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
        107110    107120    107130    107140    107150    107160    107170

      130       140         150         160         170         180         190
    CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
    | |   | ||  ||     |    | |  | | ||  || | |  ||   ||  |  || || | | | | ||
```

```
CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
      107180    107190    107200    107210    107220    107230

   200       210       220       230       240       250       260
-TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
 GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
107240    107250    107260    107270    107280    107290    107300

        270       280       290       300       310       320
CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
107310    107320    107330    107340    107350    107360    107370

  330       340       350       360       370       380       390
CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
  GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
 107380    107390    107400    107410    107420    107430    107440

  400       410       420       430       440       450       460
GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
  107450    107460    107470    107480    107490    107500    107510

        470       480       490       500       510       520
GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
  107520    107530    107540    107550    107560    107570    107580

530       540       550       560       570       580       590       600
CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
107590    107600       107610    107620    107630    107640    107650

        610       620       630       640       650       660
AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
  107660    107670    107680    107690       107700    107710    107720

 670       680       690       700       710       720       730
AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
  107730    107740    107750    107760    107770    107780    107790

     740       750       760       770       780       790
AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA-----CTGGGTAATCTT
AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
107800    107810    107820    107830       107840    107850    107860

800       810       820       830       840       850       860
GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
```



```
GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
     107870    107880       107890    107900    107910    107920

 870       880       890       900       910       920       930
AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
07930     107940    107950       107960    107970    107980    107990

 940       950       960       970       980       X
GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
108000    108010    108020    108030    108040    108050
```

11. US-08-162-407-5' (1-988)
EBV Epstein-Barr virus (EBV) genome. The complete sequ

```
ID   EBV        standard; circular DNA; VRL; 172281 BP.
XX
AC   V01555; J02070; K01729; K01730; V01554; X00498; X00499; X00784;
XX
DT   06-DEC-1983 (Rel. 04, Created)
DT   04-JAN-1995 (Rel. 42, Last updated, Version 51)
XX
DE   Epstein-Barr virus (EBV) genome. The complete sequence [1-10] was
DE   determined from DNA from B95-8 cells cloned by Arrand et al [11].
DE   B95-8 is a productive marmoset lymphoblastoid cell line
DE   immortalized with human EBV from a mononucleosis patient.
XX
KW   DNA polymerase; EBNA; genome; ribonucleotide reductase;
KW   tandem repeat; terminal repeat.
XX
OS   Epstein-Barr virus (EBV)
OC   Viridae; ds-DNA enveloped viruses; Herpesviridae;
OC   Gammaherpesviridae.
XX
RN   [1]
RP   1-172281
RA   Baer R.J., Bankier A.T., Biggin M.D., Deininger P.L.,
RA   Farrell P.J., Gibson T.J., Hatfull G.F., Hudson G.S.,
RA   Satchwell S.C., Seguin C., Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus
RT   genome";
RL   Nature 310:207-211(1984).
XX
RN   [2]
RA   Deininger P.L., Bankier A.T., Farrell P.J., Baer R., Barrell B.G.;
RT   "Sequence analysis and in vitro transcription of portions of the
RT   Epstein-Barr virus genome";
RL   J. Cell. Biochem. 19:267-274(1982).
XX
RN   [3]
RA   Farrell P.J., Deininger P.L., Bankier A., Barrell B.;
RT   "Homologous upstream sequences near Epstein-Barr virus promoters";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:1565-1569(1983).
XX
RN   [4]
RA   Farrell P.J., Bankier A.T., Seguin C., Deininger P.L.,
```

```
RA   Barrell B.G.;
RT   "Latent and lytic cycle promoters of Epstein-Barr virus";
RL   EMBO J. 2:1331-1338(1983).
XX
RN   [5]
RP   142687-159853
RA   Bankier A.T., Deininger P.L., Farrell P.J., Barrell B.G.;
RT   "Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8
RT   Epstein-Barr virus";
RL   Mol. Biol. Med. 1:21-45(1983).
XX
RN   [6]
RP   112620-125316
RA   Seguin C., Farrell P.J., Barrell B.G.;
RT   "DNA sequence and transcription of the BamHI fragment B region of
RT   B95-8 Epstein-Barr virus";
RL   Mol. Biol. Med. 1:369-392(1983).
XX
RN   [7]
RP   159853-172281
RA   Bankier A.T., Deininger P.L., Satchwell S.C., Baer R.,
RA   Farrell P.J., Barrell B.G.;
RT   "DNA sequence analysis of the EcoRI Dhet fragment of B95-8
RT   Epstein-Barr virus containing the terminal repeat sequences";
RL   Mol. Biol. Med. 1:425-446(1983).
XX
RN   [8]
RP   87650-92703
RA   Biggin M.D., Farrell P.J., Barrell B.G.;
RT   "Transcription and DNA sequence analysis of the BamHI L fragment
RT   of B95-8 Epstein-Barr virus";
RL   EMBO J. 3:1083-1090(1984).
XX
RN   [9]
RP   76089-79808
RA   Gibson T.J., Stockwell P., Ginsburg M., Barrell B.G.;
RT   "Homology between two EBV early genes and HSV ribonucleotide
RT   reductase and 38K genes";
RL   Nucleic Acids Res. 12:5087-5099(1984).
XX
RN   [10]
RA   Hatfull G.F., Barrell B.G., Quinn J., McGeoch D.;
RT   ;
RL   Unpublished.
XX
RN   [11]
RA   Arrand J.R., Walsh J.E., Rymo L., Bjoerck E., Lindahl T.,
RA   Griffin B.E.;
RT   "Molecular cloning of the complete Epstein-Barr virus genome as a
RT   set of overlapping restriction endonuclease fragments";
RL   Nucleic Acids Res. 9:2999-3014(1981).
XX
RN   [12]
RA   Kozak M.;
RT   "Possible role of flanking nucleotides in recognition of the AUG
RT   initiator codon by eukaryotic ribosomes";
RL   Nucleic Acids Res. 9:5233-5252(1981).
XX
RN   [13]
RP   7315-9312
```



```
RA   Yates J., Warren N., Reisman D., Sugden B.;
RT   "A cis-acting element from the Epstein-Barr viral genome that
RT   permits stable replication of recombinant plasmids";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:3806-3810(1984).
XX
RN   [14]
RP   45415-52824
RA   Jones M.D., Foster L., Sheedy T., Griffin B.E.;
RT   "The EB virus genome in Daudi Burkitt's lymphoma cells has a
RT   deletion similar to that observed in a non transforming strain
RT   (P3HR-1) of the virus";
RL   EMBO J. 3:813-821(1984).
XX
RN   [15]
RP   45644-52450
RA   Jeang K.T., Hayward S.D.;
RT   "Organisation of the Epstein-Barr virus molecule. III. The
RT   location of the P3HR-1 deletion junction and characterization of
RT   the NotI repeat units that form part of the template for an
RT   abundant 12-O-Tetradecanoylphorbol-13-acetate-induced mRNA
RT   transcript";
RL   J. Virol. 48:135-148(1983).
XX
RN   [16]
RP   1-172281
RA   Farrell P.J., Barrell B.G.;
RT   ;
RL   Submitted (05-JUN-1984) to the EMBL/GenBank/DDBJ databases.
XX
RN   [17]
RA   Bodescot M., Perricaudet M.;
RT   "Clustered alternative splice sites in Epstein-Barr virus RNAs";
RL   Nucleic Acids Res. 15:5887-5887(1987).
XX
RN   [18]
RA   Laux G., Perricaudet M., Farrell P.J.;
RT   "A spliced Epstein-Barr virus gene expressed in immortalized
RT   lymphocytes is created by circularization of the linear viral
RT   genome";
RL   EMBO J. 7:769-774(1988).
XX
RN   [19]
RP   1-172281
RA   Farrell P.J.;
RT   ;
RL   Submitted (18-MAR-1988) to the EMBL/GenBank/DDBJ databases.
RL   Farrell P., Ludwig Institute for Cancer Research, St. Mary's
RL   Hospital Medical School, Norfolk Place London W2 1PG.
XX
DR   EPD; 07138; EBV[DL/R].
DR   EPD; 07139; EBV[BamH1-L] R1.
DR   EPD; 07140; EBV[BamH1-L] R2.
DR   EPD; 07141; EBV[BamH1-L] L2.
DR   EPD; 07142; EBV[BamH1-L] L1.
DR   EPD; 07143; EBV[BamH1-L] L3.
DR   EPD; 07144; EBV[BamH1-K] 2.1 kb.
DR   EPD; 07145; EBV[BamH1-K] 1.3 kb.
DR   EPD; 07146; EBV[EcoR1-H] L1.
DR   EPD; 07147; EBV[EcoR1-C] L1.
DR   EPD; 07148; EBV[EcoR1-D] L1.
```

```
DR   EPD; 11192; EBV[EcoR1-D] R1.
DR   EPD; 11193; EBV[EcoR1-D] L2.
DR   EPD; 11194; EBV[EcoR1-D] L1A.
DR   EPD; 16087; EBV[BamHI-N] R1 145K.
DR   EPD; 16088; EBV[BamHI-C] R1 20K.
DR   EPD; 16089; EBV[BamHI-F] R1.
DR   EPD; 16090; EBV[BamHI-F] R2.
DR   EPD; 16091; EBV[BamHI-F] L2.
DR   EPD; 26011; EBV EBNA mRNAs   E1.
DR   EPD; 26012; EBV EBNA mRNAs   E2.
DR   EPD; 26014; EBV[BamH1-Z] L1 1 kb.
DR   SWISS-PROT; P03175; RIR2_EBV.
DR   SWISS-PROT; P03177; KITH_EBV.
DR   SWISS-PROT; P03179; VP14_EBV.
DR   SWISS-PROT; P03180; BCRF_EBV.
DR   SWISS-PROT; P03181; YHL1_EBV.
DR   SWISS-PROT; P03182; EAR_EBV.
DR   SWISS-PROT; P03183; UL31_EBV.
DR   SWISS-PROT; P03184; UL32_EBV.
DR   SWISS-PROT; P03185; UL34_EBV.
DR   SWISS-PROT; P03186; TEGU_EBV.
DR   SWISS-PROT; P03187; VP19_EBV.
DR   SWISS-PROT; P03188; VGLB_EBV.
DR   SWISS-PROT; P03189; UL37_EBV.
DR   SWISS-PROT; P03190; RIR1_EBV.
DR   SWISS-PROT; P03191; EAD_EBV.
DR   SWISS-PROT; P03192; YMR2_EBV.
DR   SWISS-PROT; P03193; UL52_EBV.
DR   SWISS-PROT; P03194; YSR1_EBV.
DR   SWISS-PROT; P03195; DUT_EBV.
DR   SWISS-PROT; P03196; UL73_EBV.
DR   SWISS-PROT; P03197; YLR2_EBV.
DR   SWISS-PROT; P03198; DPOL_EBV.
DR   SWISS-PROT; P03199; YLL2_EBV.
DR   SWISS-PROT; P03200; VGP3_EBV.
DR   SWISS-PROT; P03201; VGP2_EBV.
DR   SWISS-PROT; P03202; YLR3_EBV.
DR   SWISS-PROT; P03203; EBN4_EBV.
DR   SWISS-PROT; P03204; EBN6_EBV.
DR   SWISS-PROT; P03205; YZL2_EBV.
DR   SWISS-PROT; P03206; BZLF_EBV.
DR   SWISS-PROT; P03207; YRR1_EBV.
DR   SWISS-PROT; P03208; VMEM_EBV.
DR   SWISS-PROT; P03209; BRL1_EBV.
DR   SWISS-PROT; P03210; YRR2_EBV.
DR   SWISS-PROT; P03211; EBN1_EBV.
DR   SWISS-PROT; P03212; YKR2_EBV.
DR   SWISS-PROT; P03213; UL06_EBV.
DR   SWISS-PROT; P03214; HELI_EBV.
DR   SWISS-PROT; P03215; VIMP_EBV.
DR   SWISS-PROT; P03216; UL11_EBV.
DR   SWISS-PROT; P03217; EXON_EBV.
DR   SWISS-PROT; P03218; VGL2_EBV.
DR   SWISS-PROT; P03219; VTER_EBV.
DR   SWISS-PROT; P03220; UL95_EBV.
DR   SWISS-PROT; P03221; UL16_EBV.
DR   SWISS-PROT; P03222; UL17_EBV.
DR   SWISS-PROT; P03223; UL92_EBV.
DR   SWISS-PROT; P03224; VGP8_EBV.
DR   SWISS-PROT; P03225; BDL2_EBV.
```



```
DR   SWISS-PROT; P03226; VCAP_EBV.
DR   SWISS-PROT; P03227; DNBI_EBV.
DR   SWISS-PROT; P03228; BRF1_EBV.
DR   SWISS-PROT; P03229; YAL1_EBV.
DR   SWISS-PROT; P03230; LMP1_EBV.
DR   SWISS-PROT; P03231; VGLH_EBV.
DR   SWISS-PROT; P03232; UL24_EBV.
DR   SWISS-PROT; P03233; UL25_EBV.
DR   SWISS-PROT; P03234; VP40_EBV.
DR   SWISS-PROT; P03235; YEC4_EBV.
DR   SWISS-PROT; P12888; UNG_EBV.
DR   SWISS-PROT; P12977; EBN3_EBV.
DR   SWISS-PROT; P12978; EBN2_EBV.
DR   SWISS-PROT; P13285; LMP2_EBV.
DR   SWISS-PROT; P13288; KR2_EBV.
DR   SWISS-PROT; P14347; UL49_EBV.
DR   SWISS-PROT; P14348; VP26_EBV.
DR   SWISS-PROT; P25214; VP23_EBV.
DR   SWISS-PROT; P25215; UL87_EBV.
DR   SWISS-PROT; P25939; IC18_EBV.
DR   SWISS-PROT; P29882; UL07_EBV.
DR   SWISS-PROT; P30117; YKR4_EBV.
DR   SWISS-PROT; P30118; YBL2_EBV.
DR   SWISS-PROT; P30119; YTR1_EBV.
DR   TRANSFAC; 00194; Release Number 1.0.
DR   TRANSFAC; 00195; Release Number 1.0.
DR   TRANSFAC; 00196; Release Number 1.0.
DR   TRANSFAC; 00197; Release Number 1.0.
DR   TRANSFAC; 00198; Release Number 1.0.
DR   TRANSFAC; 00199; Release Number 1.0.
DR   TRANSFAC; 00295; Release Number 1.0.
DR   TRANSFAC; 00296; Release Number 1.0.
DR   TRANSFAC; 00297; Release Number 1.0.
DR   TRANSFAC; 00298; Release Number 1.0.
DR   TRANSFAC; 00299; Release Number 1.0.
DR   TRANSFAC; 00300; Release Number 1.0.
DR   TRANSFAC; 00301; Release Number 1.0.
DR   TRANSFAC; 00302; Release Number 1.0.
DR   TRANSFAC; 00710; Release Number 1.0.
DR   TRANSFAC; 00923; Release Number 1.0.
DR   TRANSFAC; 02818; Release Number 1.0.
DR   TRANSFAC; 03182; Release Number 1.0.
DR   TRANSFAC; 03411; Release Number 1.0.
DR   TRANSFAC; 03412; Release Number 1.0.
XX
CC   CDS
CC
CC   Listed under this feature are all known protein coding regions as
CC   well as all the major open reading frames in the sequence. In
CC   general the term major is taken as the longest frame in a
CC   particular region taking into account the adjacent longest frames
CC   and likely transcription signals. Note that on this basis some long
CC   overlapping frames have been excluded and on the other hand some
CC   small frames have been included which might represent exons or
CC   genes because they occur in a logical combination with other
CC   features or because of some other experimental data. The reading
CC   frames are named according to the Bam H1 fragment in which they
CC   start. eg BALF3 is the third leftward frame starting in Bam H1
CC   fragment A. BORF1 is the first rightward frame in Bam H1 fragment
CC   O. If there is an obvious TATA sequence followed by an in frame Met
```

```
CC   codon that satisfies the rules of Kozak [12] in that there is a
CC   purine at -3 and/or a G at +4 then the reading frame is numbered
CC   from the A of the ATG to the base preceding the termination codon.
CC   If there is no obvious initiation codon or there is a substantial
CC   reading frame in phase before the ATG then the reading frame is
CC   numbered from the first base of the first codon.
CC
CC
CC   SITEs of POLYA signals
CC
CC   This feature lists all occurences of the sequence AATAAA which is
CC   found normally approximately 20 bases upstream of the mRNA
CC   processing/polyA addition site. The rarely used homolog ATTAAA is
CC   only listed when it is found in a position close to the end of a
CC   major reading frame.
CC
CC
CC   SITEs of DONOR and ACCEPT sequences
CC
CC   This is not a comprehensive listing of all such sequences and only
CC   the positions of a few have been noted because they occur in
CC   potentially interesting positions. The number quoted in the table
CC   is the position of the terminal base in the intron in each case.
CC
CC
CC   Restriction enzyme SITEs.
CC
CC   Only the positions of the sites Bam HI (BAM) are listed.
CC
CC
CC   RPT
CC
CC   This feature is used to define repetitive sequences.
CC
CC
CC   SITE DEL
CC
CC   This feature defines deletions in B95-8 with respect to other
CC   strains such as RAJI and also to deletions in other strains such as
CC   P3HR1 and DAUDI with respect to B95-8.
CC
CC
CC   SITE HPN
CC
CC   Denotes sequences with twofold symmetry ie could form hairpin
CC   loops. This is not a comprehensive list - only a few occurences
CC   noted.
CC
CC
CC   ORGRPL
CC
CC   Denotes the region that encompasses an origin of replication
CC   (ori P).[13].
CC
CC
CC   NUMBERING
CC
CC   The DNA sequence of B95-8 EBV has been revised [19]. The original
CC   (Baer et al, 1984) base 359 has been deleted so the new sequence
CC   around that position reads TCAGTCTTT. To avoid renumbering the
```



```
CC   entire sequence, position 1 has benn moved 1 base to the left of
CC   the EcoRI site separating EcoRI Dhet from EcoRI I
CC   (ie the first A of AGAATTC)
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..172281
FT                   /organism="Epstein-Barr virus"
FT                   /strain="B95-8"
FT   mRNA            58..272
FT                   /note="exon 2 terminal protein RNA"
FT   mRNA            360..458
FT                   /note="exon 3 terminal protein RNA"
FT   misc_feature    complement(535..535)
FT                   /note="polyA signal: AATAAA"
FT   mRNA            540..788
FT                   /note="exon 4 terminal protein RNA"
FT   mRNA            871..951
FT                   /note="exon 5 terminal protein RNA"
FT   mRNA            1026..1196
FT                   /note="exon 6 terminal protein RNA"
FT   promoter        complement(1192..1192)
FT                   /note="TATA: TATAAAT"
FT   mRNA            1280..1495
FT                   /note="exon 7 terminal protein RNA"
FT   promoter        complement(1383..1383)
FT                   /note="TATA: CATAAAA"
FT   mRNA            1574..1682
FT                   /note="exon 8 terminal protein RNA"
FT   promoter        1676..1676
FT                   /note="TATA: TATAAAG"
FT   promoter        1691..1691
FT                   /note="TATA: TATTAAA BN-R1 late promoter before BNRF1,
FT                   gives 4.1kb late RNA. Probably encodes non glycosylated
FT                   140kd protein in membrane antigen. Also two latent RNAs
FT                   spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson
FT                   et al, 1985). The longer one encodes terminal protein."
FT   CDS             1736..5692
FT                   /note="BNRF1 reading frame, 5 NXT/S"
FT   misc_feature    complement(1795..1795)
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    3955..3955
FT                   /note="BAM: Bam H1 Nhet/h"
FT   misc_feature    3994..3994
FT                   /note="BAM: Bam H1 h/C"
FT   mRNA            5408..5856
FT                   /note="exon 9 terminal protein RNA"
FT   misc_feature    5841..5841
FT                   /note="polyA signal: AATAAA, end of 4.1kb late RNA and TP
FT                   latent RNA."
FT   misc_feature    5863..5863
FT                   /note="alternative end to TP cDNAs"
FT   promoter        6097..6097
FT                   /note="TATA: TATAAGA"
FT   misc_feature    6629..6795
FT                   /note="Pol III RNA EBER 1"
FT   promoter        complement(6823..6823)
FT                   /note="TATA: CATAAAT"
FT   misc_feature    6956..7128
FT                   /note="Pol III RNA EBER 2"
```

```
FT   rep_origin      7315..9312
FT                   /note="origin of replication, ori P (Yates et al, 1984,
FT                   1985)"
FT   repeat_region   7421..8042
FT                   /note="21x30bp repeats, binding sites for EBNA-1 (site I,
FT                   Rawlins et al, 1985). Tandem repeat part of oriP (Reisman
FT                   et al, 1985). Also functions as a cell type specific
FT                   enhancer (Reisman et al, 1985; Lupton and Levine, 1985)"
FT   promoter        7738..7738
FT                   /note="TATA: TATAAAT"
FT   promoter        7888..7888
FT                   /note="TATA: TATAAAT"
FT   promoter        8573..8573
FT                   /note="TATA: CATAAAT"
FT   misc_feature    complement(8680..8680)
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    complement(8755..8755)
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    8962..8962
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    9021..9133
FT                   /note="HPN: dyad symmetry, site II for EBNA-1 binding
FT                   (Rawlins et al, 1985). Dyad symmetry part of oriP (Reisman
FT                   et al, 1985)"
FT   promoter        complement(9398..9398)
FT                   /note="TATA: TATAAAT"
FT   promoter        9631..9631
FT                   /note="TATA: TATAAAT BC-R1 late promoter before BCRF1"
FT   CDS             9675..10187
FT                   /note="BCRF1 reading frame"
FT   promoter        10076..10076
FT                   /note="TATA: GATAAAA"
FT   misc_feature    complement(10148..10148)
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    10173..10173
FT                   /note="polyA signal: ATTAAA"
FT   misc_feature    10257..10257
FT                   /note="polyA signal: AATAAA, end of 0.8kb late RNA from
FT                   BCR1 and end of 1.6 kb late RNA, start unknown"
FT   misc_feature    complement(10277..10277)
FT                   /note="polyA signal: AATAAA"
FT   promoter        complement(10975..10975)
FT                   /note="TATA: CATAAAT"
FT   promoter        11305..11305
FT                   /note="TATA: TACAAAA; BCR2 promoter for highly spliced EBN
A
FT                   latent RNAs."
FT   mRNA            11336..11480
FT                   /note="exon C1 of Bodescot et al (1986) RNAs"
FT   promoter        11524..11524
FT                   /note="TATA: TATAATT"
FT   misc_feature    complement(11587..11587)
FT                   /note="polyA signal: AATAAA"
FT   promoter        complement(11606..11606)
FT                   /note="TATA: CATAAAT"
FT   mRNA            11626..11657
FT                   /note="exon C2 of Bodescot et al (1986) RNAs"
FT   promoter        11796..11796
FT                   /note="TATA: TATAAGT"
FT   promoter        complement(11799..11799)
```



```
FT                   /note="TATA: TATAAAA"
FT   repeat_region   12001..15072
FT                   /note="3072 repeat 1"
FT   CDS             12541..13689
FT                   /note="BCRF2 3072 repeat, reading frame 1"
FT   misc_feature    13215..13215
FT                   /note="BAM: BamH1 C/W"
FT   promoter        14352..14352
FT                   /note="*TATA: TATAAAG BWR1 one of the promoters for highly
FT                   spliced EBNA and LP RNAs (Sample et al, 1986; Speck et al,
FT                   1986)"
FT   mRNA            14384..14410
FT                   /note="*exon W0 of EBNA/LP RNAs"
FT   mRNA            14554..14619
FT                   /note="*exon W1 (also W66) part of leader protein (LP)
FT                   gene. LP is also called EBNA-5 (Dillner et al, 1986) and
FT                   EBNA-4 (Rowe et al, 1987)."
FT   mRNA            14559..14619
FT                   /note="*exon W1' (also W61) of EBNA/LP RNAs forms initiato
r
FT                   met when fused to exon W0 or exon C2."
FT   mRNA            14701..14832
FT                   /note="*exon W2 (also W132) part of LP gene"
FT   repeat_region   15073..18144
FT                   /note="3072 repeat 2"
FT   CDS             15613..16761
FT                   /note="BWRF1 reading frame 2"
FT   misc_feature    16287..16287
FT                   /note="BAM: BamH1 W/W"
FT   promoter        17424..17424
FT                   /note="TATA: TATAAAG"
FT   mRNA            17626..17691
FT                   /note="Exon W1"
FT   mRNA            17773..17904
FT                   /note="Exon W2"
FT   repeat_region   18145..21216
FT                   /note="3072 repeat 3"
FT   CDS             18685..19833
FT                   /note="BWRF1 reading frame 3"
FT   misc_feature    19359..19359
FT                   /note="BAM: BamH1 W/W"
FT   promoter        20496..20496
FT                   /note="TATA: TATAAAG"
FT   mRNA            20698..20763
FT                   /note="Exon W1"
FT   mRNA            20845..20976
FT                   /note="Exon W2"
FT   repeat_region   21217..24288
FT                   /note="3072 repeat 4"
FT   CDS             21757..22905
FT                   /note="BWRF1 reading frame 4"
FT   misc_feature    22431..22431
FT                   /note="BAM: BamH1 W/W"
FT   promoter        23568..23568
FT                   /note="TATA: TATAAAG"
FT   mRNA            23771..23835
FT                   /note="Exon W1"
FT   mRNA            23917..24048
FT                   /note="Exon W2"
FT   repeat_region   24289..27360
```

```
FT                   /note="3072 repeat 5"
FT   CDS             24829..25977
FT                   /note="BWRF1 reading frame 5"
FT   misc_feature    25503..25503
FT                   /note="BAM: BamH1 W/W"
FT   promoter        26640..26640
FT                   /note="TATA: TATAAAG"
FT   mRNA            26842..26907
FT                   /note="Exon W1"
FT   mRNA            26989..27120
FT                   /note="Exon W2"
FT   repeat_region   27361..30432
FT                   /note="3072 repeat 6"
FT   CDS             27901..29049
FT                   /note="BWRF1 reading frame 6"
FT   misc_feature    28575..28575
FT                   /note="BAM: BamH1 W/W"
FT   promoter        29712..29712
FT                   /note="TATA: TATAAAG"
FT   mRNA            29914..29979
FT                   /note="Exon W1"
FT   mRNA            30061..30192
FT                   /note="Exon W2"
FT   repeat_region   30433..33504
FT                   /note="3072 repeat 7"
FT   CDS             30973..32121
FT                   /note="BWRF1 reading frame 7"
FT   misc_feature    31647..31647
FT                   /note="BAM: BamH1 W/W"
FT   promoter        32784..32784
FT                   /note="TATA: TATAAAG"
FT   mRNA            32986..33051
FT                   /note="Exon W1"
FT   mRNA            33133..33264
FT                   /note="Exon W2"
FT   repeat_region   33505..36576
FT                   /note="3072 repeat 8"
FT   CDS             34045..35193
FT                   /note="BWRF1 reading frame 8"
FT   misc_feature    34719..34719
FT                   /note="BAM: BamH1 W/W"
FT   promoter        35856..35856
FT                   /note="TATA: TATAAAG"
FT   mRNA            36058..36123
FT                   /note="Exon W1"
FT   mRNA            36205..36336
FT                   /note="Exon W2"
FT   repeat_region   36577..39648
FT                   /note="3072 repeat 9"
FT   CDS             37117..38265
FT                   /note="BWRF1 reading frame 9"
FT   misc_feature    37791..37791
FT                   /note="BAM: BamH1 W/W"
FT   promoter        38928..38928
FT                   /note="TATA: TATAAAG"
FT   mRNA            39130..39195
FT                   /note="Exon W1"
FT   mRNA            39277..39408
FT                   /note="Exon W2"
FT   repeat_region   39649..42720
```



```
FT                   /note="3072 repeat 10"
FT   CDS             40189..41337
FT                   /note="BWRF1 reading frame 10"
FT   misc_feature    40863..40863
FT                   /note="BAM: BamH1 W/W"
FT   promoter        42000..42000
FT                   /note="TATA: TATAAAG"
FT   mRNA            42202..42267
FT                   /note="Exon W1"
FT   mRNA            42349..42480
FT                   /note="Exon W2"
FT   repeat_region   42721..45792
FT                   /note="3072 repeat 11"
FT   CDS             43261..44409
FT                   /note="BWRF1 reading frame 11"
FT   misc_feature    43935..43935
FT                   /note="BAM: BamH1 W/W"
FT   promoter        45072..45072
FT                   /note="TATA: TATAAAG"
FT   mRNA            45274..45339
FT                   /note="Exon W1"
FT   misc_feature    45415..52824
FT                   /note="DEL: DAUDI deletion (Jones et al, 1984)"
FT   mRNA            45421..45552
FT                   /note="Exon W2"
FT   misc_feature    45644..52450
FT                   /note="DEL: P3HR1 deletion (Jeang and Hayward, 1983)"
FT   repeat_region   45793..47643
FT                   /note="3072 repeat 12"
FT   CDS             46333..47481
FT                   /note="BWRF1 reading frame 12"
FT   misc_feature    47007..47007
FT                   /note="BAM: BamH1 W/Y"
FT   mRNA            47761..47793
FT                   /note="Exon Y1 Bodescot et al, 1984"
FT   promoter        47831..47831
FT                   /note="TATA: TATAAGT"
FT   mRNA            47878..47999
FT                   /note="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and
FT                   Strominger,1985), last common exon"
FT   misc_feature    complement(48023..48023)
FT                   /note="polyA signal: AATAAA"
FT   CDS             48386..50032
FT                   /note="Coding exon for EBNA-2 (Sample et al,1986)"
FT   mRNA            48386..48444
FT                   /note="exon Bodescot et al, 1984"
FT   CDS             48429..49964
FT                   /note="BYRF1, encodes EBNA-2 (Dambaugh et al, 1984; Dillne
r
FT                   et al, 1984)"
FT   repeat_region   48678..48800
FT                   /note="14 x CCCCCACCA repeats"
FT   misc_feature    48848..48848
FT                   /note="BAM: BamH1 Y/H"
FT   promoter        49350..49350
FT                   /note="TATA: TATAACA"
FT   promoter        complement(49353..49353)
FT                   /note="TATA: TATAAAA"
FT   repeat_region   49525..49578
FT                   /note="9 x GGGGCA repeats"
```

```
FT   mRNA            49852..50032
FT                   /note="exon (Bodescot et al 1984)"
FT   misc_feature    50003..50003
FT                   /note="polyA signal: AATAAA, end of Bodescot T1 RNA and
FT                   EBNA-2 RNA (3.0kb latent RNA in IB4 cells)"
FT   promoter        complement(50156..50156)
FT                   /note="TATA: TATAAGT"
FT   misc_feature    complement(50317..50317)
FT                   /note="polyA signal: AATAAA, end of 2.5kb early RNA from
FT                   52817"
FT   repeat_region   50578..52115
FT                   /note="12 x 125bp repeats"
FT   misc_feature    complement(50578..52557)
FT                   /note="BHLF1 early reading frame"
FT   misc_feature    52654..53697
FT                   /note="region homologous to Eco R1 C of Raji"
FT   promoter        complement(52817..52817)
FT                   /note="TATA: GATAAAA promoter for 2.5kb early RNA
FT                   containing BHLF1 (Jeang and Hayward, 1983; Freese et
FT                   al,1983)"
FT   promoter        53759..53759
FT                   /note="TATA: TATTAAC likely promoter for class III and IV
FT                   early RNAs encoding BHRF1 (Pearson et al, 1987)"
FT   misc_feature    53895..53895
FT                   /note="DONOR: CGGGTAACT donor for splice to 54335 in class
FT                   IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
FT   misc_feature    54335..54335
FT                   /note="ACCEPT: TTTTCTAG acceptor from 48444 in class I,
FT                   47999 in class II, and 53895 in class IV early RNAs
FT                   encoding BHRF1 (Pearson et al, 1987)"
FT   misc_feature    54376..54948
FT                   /note="BHRF1 reading frame, limited homolgy to bcl-2 gene.
FT                   Early gene in B95-8 cells and part of restricted EA
FT                   complex."
FT   promoter        54591..54591
FT                   /note="TATA: TATAACA"
FT   promoter        complement(54594..54594)
FT                   /note="TATA: TATAAAT"
FT   misc_feature    54853..54853
FT                   /note="BAM: BamH1 H/F"
FT   misc_feature    complement(54929..54929)
FT                   /note="polyA signal: AATAAA"
FT   promoter        complement(54977..54977)
FT                   /note="TATA: TATAAAG"
FT   misc_feature    55518..55518
FT                   /note="polyA signal: AATAAA, 3' end of 2.5kb, 1.9kb, 1.7kb
FT                   and 0.6kb early RNAs"
FT   misc_feature    complement(55990..55990)
FT                   /note="polyA signal: AATAAA, 3' end of 2.3kb and 1.1kb
FT                   early RNAs from 58568 and 57081"
FT   promoter        complement(56132..56132)
FT                   /note="TATA: TATAAAG"
FT   misc_feature    complement(55982..56935)
FT                   /note="BFLF2 reading frame, 4 NXT/S, homologous to RF 27 i
n
FT                   VZV and HFRF2 in CMV"
FT   promoter        complement(57081..57081)
FT                   /note="TATA: TATTTAA before BFLF2; BFL2 promoter gives
FT                   1.1kb early RNA"
FT   promoter        complement(58088..58088)
```



```
FT                   /note="TATA: GATAAAA"
FT   CDS             complement(56951..58525)
FT                   /note="BFLF1 reading frame, 2 NXT/S homologous to RF 26 in
FT                   VZV and HFRF1 in CMV"
FT   promoter        complement(58568..58568)
FT                   /note="TATA: TATTAAA before BFLF1, BFL1 promoter gives
FT                   2.3kb early RNA"
FT   promoter        58832..58832
FT                   /note="TATA: TATAAAA before BFRF1"
FT   CDS             58891..59901
FT                   /note="BFRF1 early reading frame, 1 NXT/S, homologous to
FT                   HFLF4 in CMV"
FT   CDS             59610..61580
FT                   /note="BFRF2 early reading frame, homologous to HFLF5 in
FT                   CMV"
FT   promoter        complement(61062..61062)
FT                   /note="TATA: GATAAAA"
FT   promoter        61344..61344
FT                   /note="TATA: TATTTAA before BFRF3"
FT   CDS             61456..62034
FT                   /note="BFRF3 early reading frame"
FT   misc_feature    complement(62068..62068)
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    62069..62069
FT                   /note="polyA signal: AATAAA, 3' end of 10, 6.5, 3.7, 3.4,
FT                   3.1, 2.5 and 0.8kb early RNAs"
FT   misc_feature    62249..62249
FT                   /note="BAM: BamH1 F/Q"
FT   misc_feature    62430..62477
FT                   /note="Site III for EBNA-1 binding (Rawlins et al, 1985)"
FT   misc_feature    66121..66121
FT                   /note="BAM: BamH1 Q/U"
FT   mRNA            67477..67649
FT                   /note="Exon in EBNA-1 RNA (Speck and Strominger,1985) and
FT                   cDNA clone T4 (Bodescot et al, 1986)"
FT   misc_feature    69410..69410
FT                   /note="BAM: BamH1 U/P"
FT   repeat_region   69684..69930
FT                   /note="5 x 51bp repeats"
FT   repeat_region   70387..70521
FT                   /note="9 x 15bp repeat"
FT   promoter        70750..70750
FT                   /note="TATA: CATAAAA"
FT   CDS             complement(62081..71527)
FT                   /note="BPLF1 reading frame, 1 NXT/S, analogous to VZV RF22
"
FT   promoter        complement(72192..72192)
FT                   /note="TATA: TATTAAA before BPLF1"
FT   misc_feature    73468..73468
FT                   /note="BAM: BamH1 P/O"
FT   promoter        75017..75017
FT                   /note="TATA: TATTTAA BO-R1 late promoter before BORF1,
FT                   gives 3.9kb late RNA"
FT   CDS             75238..76332
FT                   /note="BORF1 late reading frame, 2 NXT/S homologous to VZV
FT                   RF20"
FT   CDS             complement(71523..75239)
FT                   /note="BOLF1 reading frame, 1 NXT/S analogous to VZV RF 21
"
FT   promoter        complement(75322..75322)
```

```
FT                   /note="TATA: TATTTAG before BOLF1"
FT   promoter        75819..75819
FT                   /note="TATA: TATAAAG"
FT   misc_feature    75838..75838
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    complement(76126..76126)
FT                   /note="polyA signal: AATAAA"
FT   promoter        76169..76169
FT                   /note="TATA: TACATAT BO-R2 early promoter before BORF2,
FT                   gives 2.8kb RNA"
FT   misc_feature    complement(76300..76300)
FT                   /note="polyA signal: AATAAA"
FT   CDS             76407..78887
FT                   /note="BORF2 early reading frame, 2 NXT/S. Homology HSV
FT                   140K ribonucleotide reductase (Gibson et al, 1984) and RF
FT                   19 VZV"
FT   misc_feature    77835..77835
FT                   /note="BAM: Bam H1 O/a"
FT   promoter        78804..78804
FT                   /note="TATA: TATAAGT Ba-R1 early promoter before BaRF1,
FT                   gives 3.5kb RNA"
FT   misc_feature    78883..78883
FT                   /note="polyA signal: AATAAA, end of 3.9kb late RNA from
FT                   75017 and 2.8kb early RNA from 76169"
FT   misc_feature    complement(78896..78896)
FT                   /note="polyA signal: AATAAA"
FT   CDS             78900..79808
FT                   /note="BaRF1 early reading frame, homologous to HSV 38K
FT                   ribonucleotide reductase (Gibson et al, 1984) and RF 18
FT                   VZV"
FT   promoter        complement(79495..79495)
FT                   /note="TATA: TATAACA"
FT   misc_feature    79537..79537
FT                   /note="BAM: Bam H1 a/M"
FT   promoter        79840..79840
FT                   /note="TATA: CATAAAT BM-R1 early promoter before BMRF1,
FT                   gives 2.5kb RNA"
FT   CDS             79899..81113
FT                   /note="BMRF1 early reading frame. Early antigen protein
FT                   recognised by R3 monoclonal (Pearson et al 1983; Cho et al
,
FT                   1985a)"
FT   promoter        80779..80779
FT                   /note="TATA: TATTTAA BM-R2 late promoter before BMRF2"
FT   misc_feature    complement(80782..80782)
FT                   /note="polyA signal: AATAAA"
FT   promoter        80832..80832
FT                   /note="TATA: GATAAAA, possible promoter for 1.4kb late RNA
FT                   encoding BMRF2"
FT   CDS             81118..82191
FT                   /note="BMRF2 early reading frame"
FT   promoter        81751..81751
FT                   /note="TATA: CATAAAT"
FT   misc_feature    82180..82180
FT                   /note="polyA signal: ATTAAA, end of 3.5kb early RNA from
FT                   78804, 2.5kb early RNA from 79840 and 1.4kb late RNA"
FT   promoter        complement(82311..82311)
FT                   /note="TATA: CATAAAT"
FT   repeat_region   82319..82461
FT                   /note="2x71bp repeats"
```



```
FT   misc_feature    complement(82747..82747)
FT                   /note="polyA signal: AATAAA"
FT   repeat_region   83640..83729
FT                   /note="10x9bp repeats"
FT   misc_feature    complement(84122..84122)
FT                   /note="ACCEPT: CTCCCCTCTGCAG acceptor in spliced form of
FT                   BMLF1 RNA"
FT   CDS             complement(82746..84122)
FT                   /note="BMLF1 early reading frame. Diffuse early antigen
FT                   (Cho et al, 1985b). Also homologous to RF 4 VZV and IE63 o
f
FT                   HSV.(BSLF2 + BMLF1) is also called EB2 (Chevallier-Greco e
t
FT                   al, 1986). General transactivator of transcription
FT                   (Lieberman et al, 1986)."
FT   misc_feature    complement(84227..84227)
FT                   /note="DONOR: CAGGTAAGA donor in spliced form of BMLF1 RNA
"
FT   misc_feature    84233..84233
FT                   /note="BAM: Bam H1 M/S"
FT   CDS             complement(84229..84288)
FT                   /note="BSLF2 early reading frame in 5' exon of spliced RNA
FT                   encoding BMLF1"
FT   promoter        complement(84356..84356)
FT                   /note="TATA: CATAAAT before BSLF2 and BMLF1. Two RNAs star
t
FT                   here; one is spliced and the other is unspliced, both
FT                   traverse BMLF1."
FT   CDS             complement(84260..86881)
FT                   /note="BSLF1 reading frame, homologous to RF 6 VZV"
FT   promoter        86882..86882
FT                   /note="TATA: TATTTAA BS-R1 late promoter before BSRF1"
FT   CDS             86924..87580
FT                   /note="BSRF1 reading frame"
FT   misc_feature    complement(87134..87134)
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    87599..87599
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    complement(87613..87613)
FT                   /note="polyA signal: AATAAA, end 1.0kb early RNA from BLL3
"
FT   misc_feature    87650..87650
FT                   /note="BAM: Bam H1 S/L"
FT   CDS             complement(87641..88474)
FT                   /note="BLLF3 early reading frame (BLLF2 in Baer et al,
FT                   1984). Homologous to RF 8 VZV and dUTPase HSV."
FT   promoter        88507..88507
FT                   /note="TATA: TATATAT BL-R1 late promoter before BLRF1,
FT                   gives 1.0kb late RNA"
FT   promoter        88511..88511
FT                   /note="TATA: TATAAGA"
FT   promoter        complement(88514..88514)
FT                   /note="TATA: TATATAT BL-L3 early promoter before BLLF3,
FT                   gives 1.0kb early RNA"
FT   CDS             88547..88855
FT                   /note="BLRF1 late reading frame"
FT   promoter        88863..88863
FT                   /note="TATA: TATTTAA BL-R2 late promoter before BLRF2,
FT                   gives 0.6kb late RNA"
FT   CDS             88925..89413
```

```
FT                   /note="BLRF2 late reading frame, 2 NXS/T"
FT   misc_feature    89412..89412
FT                   /note="polyA signal: AATAAA, end of 1.0kb and 0.6kb late
FT                   RNAs"
FT   misc_feature    complement(89425..89425)
FT                   /note="polyA signal: AATAAA, end of 0.7kb early, 2.2kb lat
e
FT                   and 2.8kb late RNA"
FT   promoter        complement(89434..89434)
FT                   /note="TATA: TATAAAG"
FT   CDS             complement(89569..90013)
FT                   /note="BLLF2 early reading frame (BLLF3 in Baer et al,
FT                   1984)"
FT   promoter        complement(90051..90051)
FT                   /note="TATA: TATAACA BL-L2 early promoter before BLLF2,
FT                   gives 0.7kb early RNA"
FT   repeat_region   90177..90639
FT                   /note="21 copies of 21bp approximate repeat"
FT   intron          complement(90062..90652)
FT                   /note="intervening sequence in gp220 gene"
FT   CDS             complement(89433..92153)
FT                   /note="BLLF1b, late reading frame gp220 membrane antigen,
FT                   spliced form of BLLF1a (Hummel et al, 1984; Biggin et al,
FT                   1984; Beisel et al, 1985)"
FT   CDS             complement(89433..92153)
FT                   /note="BLLF1a, late reading frame, gp350 membrane antigen,
FT                   36 NXT/S (Hummel et al, 1984; Biggin et al, 1984; Beisel e
t
FT                   al, 1985)"
FT   promoter        complement(92192..92192)
FT                   /note="TATA: TATTAAA BL-L1 late promoter before BLLF1a,b.
FT                   Gives 2.8 and 2.2kb late RNAs"
FT   mRNA            92238..92581
FT                   /note="Exon in Bodescot et al (1986) RNA (spliced from
FT                   20763 to 92670)"
FT   CDS             92243..92581
FT                   /note="BLRF3 reading frame"
FT   mRNA            92670..95248
FT                   /note="Exon in (Bodescot et al, 1986) RNA from 92581, to 3
'
FT                   end"
FT   CDS             92670..95162
FT                   /note="BERF1 frame, homology with BERF2b and BERF4. A
FT                   fusion of BLRF3 with BERF1 encodes EBNA-3A, latent cycle
FT                   gene. (Hennessy et al, 1986, Joab et al, 1987)"
FT   misc_feature    92703..92703
FT                   /note="BAM: Bam H1 L/E"
FT   promoter        complement(93161..93161)
FT                   /note="TATA: CATAAAT"
FT   promoter        93479..93479
FT                   /note="TATA: TATAAGA"
FT   promoter        complement(93482..93482)
FT                   /note="TATA: TATAAAT"
FT   repeat_region   94208..94277
FT                   /note="repeat type A"
FT   repeat_region   94281..94306
FT                   /note="repeat type B"
FT   repeat_region   94307..94381
FT                   /note="repeat type C"
FT   repeat_region   94386..94411
```



```
FT                   /note="repeat type B"
FT   repeat_region   94412..94489
FT                   /note="repeat type C"
FT   repeat_region   94490..94560
FT                   /note="repeat type A"
FT   repeat_region   94571..94648
FT                   /note="repeat type C"
FT   repeat_region   94649..94719
FT                   /note="repeat type A"
FT   repeat_region   94896..94982
FT                   /note="repeat type D"
FT   repeat_region   94983..95069
FT                   /note="repeat type D"
FT   misc_feature    95221..95221
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    complement(95272..95272)
FT                   /note="polyA signal: AATAAA"
FT   CDS             95353..95724
FT                   /note="BERF2a reading frame"
FT   CDS             95725..98244
FT                   /note="BERF2b frame, homology with BERF1 and BERF4. BERF2a
FT                   and BERF2b are spliced together to make EBNA3B (EBNA4A)
FT                   latent protein."
FT   misc_feature    complement(95819..95819)
FT                   /note="polyA signal: AATAAA"
FT   promoter        complement(95853..95853)
FT                   /note="TATA: TATAAAT"
FT   misc_feature    complement(96276..96276)
FT                   /note="polyA signal: AATAAA"
FT   repeat_region   97522..97698
FT                   /note="3x60bp repeat"
FT   CDS             98323..98769
FT                   /note="BERF3 reading frame"
FT   mRNA            98364..98730
FT                   /note="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"
FT   misc_feature    98731..98731
FT                   /note="DONOR: AAGGTGAGT donor"
FT   CDS             98805..101420
FT                   /note="BERF4 frame, homology with BERF1 and BERF2b. BERF3
FT                   and BERF4 are spliced together to make the EBNA3C (EBNA 4B
)
FT                   latent protein."
FT   mRNA            98805..99050
FT                   /note="Exon in T4 cDNA (Bodescot et al 1986). 99050 is not
FT                   the end of the RNA."
FT   misc_feature    99126..102118
FT                   /note="DEL: Deletion in Raji"
FT   promoter        99443..99443
FT                   /note="TATA: CATAAAA"
FT   misc_feature    100104..100104
FT                   /note="DONOR: ACCGTGAGT possible donor before repeat."
FT   repeat_region   100122..100304
FT                   /note="10 x 15bp repeat"
FT   misc_feature    100528..100528
FT                   /note="DONOR: CTGGTAAGG possible donor"
FT   misc_feature    100613..100613
FT                   /note="BAM: Bam H1 E/e1"
FT   repeat_region   100665..100781
FT                   /note="3x39bp repeat"
FT   promoter        complement(100860..100860)
```

```
FT                   /note="TATA: TATAACA"
FT   misc_feature    100919..100919
FT                   /note="BAM: Bam H1 e1/e2"
FT   misc_feature    101426..101426
FT                   /note="BAM: Bam H1 e2/e3"
FT   promoter        101690..101690
FT                   /note="TATA: CATAAAA"
FT   misc_feature    101765..101765
FT                   /note="polyA signal: AATAAA"
FT   promoter        complement(101786..101786)
FT                   /note="TATA: TATAAAG"
FT   misc_feature    101947..101947
FT                   /note="BAM: Bam H1 e3/Z"
FT   misc_feature    complement(102098..102098)
FT                   /note="DONOR: CAGGTGAGG possible donor"
FT   CDS             complement(101448..102116)
FT                   /note="BZLF2 reading frame 3x NXT/S. 2.5kb late RNA
FT                   traverses BZLF2, ends unknown."
FT   promoter        102153..102153
FT                   /note="TATA: TATTAAT"
FT   misc_feature    complement(102156..102156)
FT                   /note="polyA signal: AATAAA 3' end of 0.9kb and 2.8kb RNAs
FT                   encoding BZLF1 and BRLF1"
FT   promoter        complement(102160..102160)
FT                   /note="TATA: TATTAAT"
FT   mRNA            complement(102126..102341)
FT                   /note="3' terminal exon of 0.9kb and 2.8kb early RNAs"
FT   promoter        complement(102380..102380)
FT                   /note="TATA: CATAAAT"
FT   promoter        102415..102415
FT                   /note="TATA:TATATAC"
FT   promoter        complement(102420..102420)
FT                   /note="TATA: TATATAC"
FT   misc_feature    complement(102504..102504)
FT                   /note="polyA signal: AATAAA, apparently not functional"
FT   mRNA            complement(102426..102530)
FT                   /note="Exon of 0.9kb and 2.8kb early RNAs"
FT   repeat_region   102581..102652
FT                   /note="semi-repetitive sequence, homologous to human c-fos
FT                   3' sequence"
FT   misc_feature    complement(102918..102918)
FT                   /note="splice acceptor used in RZ fusion gene (Sargeant)"
FT   CDS             complement(102213..103155)
FT                   /note="BZLF1 reading frame, modified from Baer et al, 1984

FT                   Has two splices within frame. 2xNXT/S. Immediate early gen
e
FT                   which disrupts latency (Countryman and Miller, 1985),
FT                   called EB1 by Chevallier-Greco et al, 1986 and ZEBRA by
FT                   Miller."
FT   mRNA            complement(102655..103194)
FT                   /note="First exon of 0.9kb early RNA encoding BZLF1"
FT   promoter        complement(103231..103231)
FT                   /note="TATA: TTTAAA of BZL1 immediate early promoter gives
FT                   0.9kb RNA"
FT   misc_feature    complement(103256..103311)
FT                   /note="Upstream of BZL1, homology to 106243 to 106188"
FT   misc_feature    complement(103453..103462)
FT                   /note="TAATGAAATC sequence"
FT   misc_feature    103741..103741
```



```
FT                   /note="BAM: Bam H1 Z/g"
FT   misc_feature    103816..103816
FT                   /note="BAM: Bam H1 g/R"
FT   mRNA            complement(104927..104989)
FT                   /note="BRLF2 poss. small 5' exon"
FT   promoter        105016..105016
FT                   /note="TATA: TATAAAT before BRRF1, possible promoter for
FT                   1.1 kb early RNA encoding BRRF1"
FT   promoter        complement(105019..105019)
FT                   /note="TATA: TATAAAT before BRLF2"
FT   CDS             105182..106114
FT                   /note="BRRF1 early reading frame"
FT   CDS             complement(103369..105183)
FT                   /note="BRLF1 reading frame, (immediate?) early gene, acts
FT                   as transcription activator."
FT   misc_feature    complement(105185..105185)
FT                   /note="ACCEPT: splice acceptor in 2.8kb early RNA encoding
FT                   BRLF1 and RZ fusion gene (Sargeant)"
FT   mRNA            complement(104926..105185)
FT                   /note="exon in RZ fusion gene (Sargeant)"
FT   promoter        105213..105213
FT                   /note="TATA: CATTAAA"
FT   misc_feature    106110..106110
FT                   /note="polyA signal: AATAAA, 3' end of early 1.1kb RNA
FT                   encoding BRRF1"
FT   misc_feature    complement(106125..106125)
FT                   /note="DONOR: CAGGTAAGA possible donor"
FT   promoter        complement(106213..106213)
FT                   /note="TATA: CATAAAA"
FT   promoter        106243..106243
FT                   /note="TATA: TATAAAA before BRRF2, possible promoter for
FT                   1.8 kb RNA encoding BRRF2"
FT   misc_feature    complement(106188..106243)
FT                   /note="Homology to upstream region of BZL1"
FT   CDS             106302..107915
FT                   /note="BRRF2 reading frame"
FT   promoter        complement(106385..106385)
FT                   /note="TATA: GATAAAA"
FT   misc_feature    complement(106973..106973)
FT                   /note="polyA signal: AATAAA"
FT   promoter        complement(107124..107124)
FT                   /note="TATA: GATAAAA"
FT   misc_feature    107457..107457
FT                   /note="BAM: Bam H1 R/f"
FT   misc_feature    107565..107565
FT                   /note="BAM: Bam H1 f/K"
FT   misc_feature    107914..107914
FT                   /note="polyA signal: AATAAA, 3' end of 1.8kb RNA encoding
FT                   BBRF2"
FT   misc_feature    107942..107942
FT                   /note="ACCEPT: splice acceptor for EBNA-1 RNA (from 98730)
"
FT   CDS             107950..109875
FT                   /note="BKRF1 encodes EBNA-1 protein, latent cycle gene."
FT   repeat_region   108217..108924
FT                   /note="EBNA triplet repeat GGA,GCA,GGG."
FT   misc_feature    109856..109856
FT                   /note="DONOR: AGGGTGAGG possible donor at end BKRF1"
FT   promoter        109905..109905
FT                   /note="TATA: TATTAAA before BKRF2, possible start for 2.3k
```

```
b
FT                   late RNA"
FT   misc_feature    109906..109906
FT                   /note="polyA signal: ATTAAA"
FT   misc_feature    109937..109937
FT                   /note="polyA signal: AATAAA 3' end of EBNA-1 RNA"
FT   CDS             109958..110371
FT                   /note="BKRF2 reading frame"
FT   misc_feature    110271..110271
FT                   /note="DONOR: TCCGTGAGT possible donor at end of BKRF2"
FT   CDS             110275..111117
FT                   /note="BKRF3 reading frame, homologous to RF 59 VZV"
FT   misc_feature    111098..111098
FT                   /note="DONOR: TCGGTGAGA possible donor at end BKRF3"
FT   CDS             111107..111784
FT                   /note="BKRF4 reading frame, contains complex repetitive
FT                   sequence"
FT   misc_feature    111272..111272
FT                   /note="DONOR: GACGTGAGT poss.donor before rpt.seq. in
FT                   BKRF4"
FT   misc_feature    111719..111719
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    111787..111787
FT                   /note="polyA signal: AATAAA : currently unknown which is 3
'
FT                   end of the 2.3kb late and 1.1kb early RNAs"
FT   misc_feature    complement(111830..111830)
FT                   /note="polyA signal: AATAAA"
FT   promoter        112471..112471
FT                   /note="TATA: TATATAT"
FT   promoter        complement(112476..112476)
FT                   /note="TATA: TATATAA"
FT   misc_feature    112620..112620
FT                   /note="BAM: Bam H1 K/B"
FT   promoter        113876..113876
FT                   /note="TATA: TATTTAT before BBRF1"
FT   promoter        complement(113885..113885)
FT                   /note="TATA: CATAAAT"
FT   CDS             114204..116045
FT                   /note="BBRF1 late reading frame, homologous to RF 54 VZV"
FT   CDS             complement(111833..114259)
FT                   /note="BBLF4 early reading frame, very good homology to
FT                   RF55 VZV"
FT   CDS             115843..116781
FT                   /note="BBRF2 late reading frame, homologous to RF 53 VZV"
FT   promoter        complement(116683..116683)
FT                   /note="TATA: GATAAAA"
FT   misc_feature    complement(116696..116696)
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    116785..116785
FT                   /note="polyA signal: AATAAA"
FT   CDS             complement(116784..117386)
FT                   /note="BBLF3 early reading frame, spliced to BBLF2. BBLF3
FT                   contains a consensus nucleotide binding site"
FT   intron          complement(117386..117515)
FT                   /note="intron spliced out in RNA linking BBLF2 and BBLF3"
FT   promoter        118981..118981
FT                   /note="TATA: TATAAAA BBR1 late promoter before BBRF3"
FT   promoter        119067..119067
FT                   /note="TATA: TTTAAAA BBR2 late promoter ?"
```



```
FT   CDS             complement(117515..119080)
FT                   /note="BBLF2 early reading frame, spliced to BBLF3"
FT   promoter        119098..119098
FT                   /note="TATA: TATTTAA BBR3 late promoter before BBRF3"
FT   misc_feature    119108..119108
FT                   /note="DONOR: AAGGTGAAT possible donor"
FT   CDS             119137..120354
FT                   /note="BBRF3 late reading frame"
FT   misc_feature    120260..120260
FT                   /note="ACCEPT: ATCTTCCTCCAGGT possible acceptor"
FT   misc_feature    120358..120358
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    complement(120764..120764)
FT                   /note="polyA signal: AATAAA, 3' end of 0.6kb late, 1.6kb
FT                   early, 3.0kb early RNAs"
FT   CDS             complement(120750..120974)
FT                   /note="BBLF1 late reading frame, possibly homologous to RF
FT                   49 VZV"
FT   promoter        complement(121331..121331)
FT                   /note="TATA: TATTAAA BBL1 late promoter before BBLF1"
FT   promoter        121669..121669
FT                   /note="TATA: CATAAAT"
FT   promoter        121697..121697
FT                   /note="TATA: TATAAAG"
FT   promoter        121772..121772
FT                   /note="TATA: CATAAAG"
FT   misc_feature    122313..122313
FT                   /note="BAM: Bam H1 B/G"
FT   CDS             complement(120932..122341)
FT                   /note="BGLF5 early reading frame, homologous to RF 48 VZV
FT                   and alkaline exonuclease of HSV"
FT   misc_feature    complement(123506..123506)
FT                   /note="DONOR: AAGGTGACT possible donor"
FT   CDS             complement(122328..123692)
FT                   /note="BGLF4 early reading frame, homologous to RF 47 VZV"
FT   promoter        complement(124117..124117)
FT                   /note="TATA: TATAAAA"
FT   misc_feature    complement(124219..124219)
FT                   /note="polyA signal: AATAAA"
FT   CDS             124938..125915
FT                   /note="BGRF1 reading frame, homologous to RF 45 VZV and
FT                   spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
FT                   Northern blots in BGRF1 detect 2.7, 2.6, 2.1kb late and
FT                   1.9kb early RNAs. 2.6, 2.1kb RNAs very weak."
FT   CDS             complement(123944..124939)
FT                   /note="BGLF3 reading frame"
FT   promoter        complement(125113..125113)
FT                   /note="TATA: TATAAAT before BGLF3"
FT   misc_feature    complement(125484..125484)
FT                   /note="polyA signal: AATAAA, 3' end of 1.6kb late, 1.8kb
FT                   late, 3.0kb late and 3.7kb early RNAs"
FT   promoter        126277..126277
FT                   /note="TATA: GATAAAA"
FT   CDS             complement(125866..126873)
FT                   /note="BGLF2 late reading frame, poor homology to RF 44
FT                   VZV"
FT   promoter        complement(126929..126929)
FT                   /note="TATA: TATTAAA EEL8 late promoter before BGLF2, give
s
FT                   1.6kb late RNA"
```

```
FT   promoter        complement(127237..127237)
FT                   /note="TATA: TATAAAA, potential promoter for 1.8kb late
FT                   RNA"
FT   misc_feature    128029..128029
FT                   /note="polyA signal: AATAAA"
FT   CDS             complement(126854..128374)
FT                   /note="BGLF1 late reading frame"
FT   promoter        complement(128432..128432)
FT                   /note="TATA: TATTTAA before BGLF1, potential promoter for
FT                   3.0kb late RNA"
FT   misc_feature    128848..128848
FT                   /note="BAM: Bam H1 G/D"
FT   CDS             complement(128347..129021)
FT                   /note="BDLF4 early reading frame"
FT   promoter        complement(129054..129054)
FT                   /note="TATA: TATTTGC before BDLF4, potential promoter for
FT                   3.7kb early RNA"
FT   CDS             129188..130348
FT                   /note="BDRF1 reading frame, homologous to RF 42 VZV and
FT                   spliced gene in HSV (Costa et al, 1985). Spliced from
FT                   BGRF1. Northern blots in BDRF1 detect 2.7, 2.6 kb late and
FT                   1.9kb early RNAs. Possibly also 1.8kb early RNA."
FT   promoter        129374..129374
FT                   /note="TATA: TATAAGC"
FT   promoter        complement(129377..129377)
FT                   /note="TATA: TATAAAG"
FT   misc_feature    129413..129413
FT                   /note="DONOR: GTGGTAAGT possible donor"
FT   misc_feature    130347..130347
FT                   /note="polyA signal: ATTAAA"
FT   misc_feature    complement(130359..130359)
FT                   /note="polyA signal: AATAAA, 3' end of 0.9kb late RNA,
FT                   2.3kb late RNA and 3.2kb late RNA"
FT   CDS             complement(130365..131066)
FT                   /note="BDLF3 late reading frame 9xNXT/S"
FT   promoter        complement(131104..131104)
FT                   /note="TATA: TATAAAA EEL4 late promoter before BDLF3, give
s
FT                   0.9kb late RNA"
FT   promoter        132266..132266
FT                   /note="TATA: TATAAAA"
FT   CDS             complement(131130..132389)
FT                   /note="BDLF2 late reading frame"
FT   promoter        complement(132476..132476)
FT                   /note="TATA: TATTTAA before BDLF2, likely promoter for
FT                   2.3kb late RNA"
FT   CDS             complement(132403..133305)
FT                   /note="BDLF1 late reading frame, poor homology to RF 41
FT                   VZV"
FT   misc_feature    complement(133312..133312)
FT                   /note="polyA signal: AATAAA, 3' end of 4.5kb late RNA"
FT   misc_feature    complement(133332..133332)
FT                   /note="DONOR: AAGGTGGTT possible donor"
FT   promoter        complement(133352..133352)
FT                   /note="TATA: TATTAAA before BDLF1"
FT   promoter        complement(133386..133386)
FT                   /note="TATA: TATATAA"
FT   misc_feature    135178..135178
FT                   /note="polyA signal: AATAAA"
FT   promoter        135394..135394
```



```
FT                   /note="TATA: TATAAGT"
FT   misc_feature    136624..136624
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    136868..136868
FT                   /note="BAM: Bam H1 D/c"
FT   CDS             complement(133324..137466)
FT                   /note="BcLF1 late reading frame, homologous to RF 40 VZV
FT                   and major capsid protein of HSV"
FT   promoter        complement(137710..137710)
FT                   /note="TATA: TATTAAA EHL1 promoter before BcLF1, gives
FT                   4.5kb late RNA"
FT   promoter        137857..137857
FT                   /note="TATA: CATAAAC"
FT   CDS             137862..139715
FT                   /note="BcRF1 reading frame"
FT   misc_feature    138019..138019
FT                   /note="BAM: Bam H1 c/b"
FT   misc_feature    139352..139352
FT                   /note="BAM: Bam H1 b/T"
FT   CDS             139642..140916
FT                   /note="BTRF1 reading frame. Northern blots detect 0.95 lat
e
FT                   and 3.8kb early RNA"
FT   misc_feature    complement(140902..140902)
FT                   /note="polyA signal: AATAAA, 3' end of 2.5kb late RNA"
FT   misc_feature    140970..140970
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    complement(141286..141286)
FT                   /note="polyA signal: AATAAA"
FT   promoter        142589..142589
FT                   /note="TATA: GATAAAA"
FT   misc_feature    142740..142740
FT                   /note="BAM: Bam H1 T/X"
FT   CDS             complement(140919..143036)
FT                   /note="BXLF2 late reading frame, encodes gp85; homologous
FT                   to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV)"
FT   promoter        complement(143310..143310)
FT                   /note="TATA: TATAAGA ECL2 late promoter before BXLF2, give
s
FT                   2.5kb late RNA"
FT   misc_feature    143608..143608
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    144791..144791
FT                   /note="ACCEPT: TCTTTCGTTTTCAGG poss. acceptor before BXRF1
"
FT   CDS             144860..145606
FT                   /note="BXRF1 late reading frame, homologous to RF 35 VZV.
FT                   Basic (core?) protein."
FT   CDS             complement(143041..144861)
FT                   /note="BXLF1 early reading frame, thymidine kinase (Little
r
FT                   et al, 1986). Weak homology to RF 36 VZV and HSV thymidine
FT                   kinase. 4.0kb early RNA presumably encodes the TK. Also a
FT                   2.2kb late RNA here."
FT   misc_feature    144862..144862
FT                   /note="BAM: Bam H1 X/V"
FT   misc_feature    144945..144945
FT                   /note="DONOR: CAGGTAAGC possible donor at 3' BXRF1"
FT   promoter        complement(145135..145135)
FT                   /note="TATA: TATAACA before BXLF1"
```

```
FT   promoter        145302..145302
FT                   /note="TATA: TATTTAA before BVRF1, potential promoter for
FT                   1.9kb early RNA"
FT   CDS             145416..147128
FT                   /note="BVRF1 early reading frame, homologous to RF 34 VZV"
FT   misc_feature    complement(146926..146926)
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    147167..147167
FT                   /note="DONOR: AAGGTAAAT possible donor"
FT   misc_feature    147170..147170
FT                   /note="polyA signal: AATAAA, 3' end of 2.4kb late and 1.9k
b
FT                   early RNAs"
FT   promoter        147721..147721
FT                   /note="TATA: TATTTAT before BVRF2, potential promoter for
FT                   2.1kb early RNA"
FT   CDS             147927..149744
FT                   /note="BVRF2 early reading frame, N-terminus homologous to
FT                   RF 33 VZV"
FT   misc_feature    148007..148007
FT                   /note="BAM: Bam H1 V/d"
FT   promoter        148620..148620
FT                   /note="TATA: TATTTAA ECR1 late promoter before BdRF1, give
s
FT                   1.2kb late RNA"
FT   CDS             148707..149741
FT                   /note="BdRF1 reading frame; this is the C terminus of
FT                   BVRF2"
FT   misc_feature    149115..149115
FT                   /note="BAM: Bam H1 d/I"
FT   misc_feature    149727..149727
FT                   /note="polyA signal: AATAAA, 3' end of 2.1kb early and
FT                   1.2kb late RNAs"
FT   misc_feature    complement(149758..149758)
FT                   /note="polyA signal: AATAAA, 3' end of 1.0kb late, 1.5kb
FT                   late and 1.8kb late RNAs"
FT   CDS             complement(149782..150525)
FT                   /note="BILF2 late reading frame 11xNXT/S"
FT   promoter        complement(150571..150571)
FT                   /note="TATA: TATTTAG before BILF2. Potential promoter for
FT                   1.0kb late RNA."
FT   repeat_region   151236..151618
FT                   /note="repetitive sequence 3X"
FT   misc_feature    151767..151767
FT                   /note="polyA signal: AATAAA"
FT   promoter        complement(151780..151780)
FT                   /note="TATA: CATAAAA"
FT   misc_feature    152012..152013
FT                   /note="DEL: B95-8 deletion with respect to Raji"
FT   promoter        152230..152230
FT                   /note="TATA: CATAAAA"
FT   CDS             complement(152164..153099)
FT                   /note="BILF1 reading frame, membrane protein, 3xNXS/T"
FT   misc_feature    153259..153259
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    153637..153637
FT                   /note="HPN: 22bp 2-fold symmetric"
FT   misc_feature    complement(153690..153690)
FT                   /note="DONOR: AAAGTGAGG possible donor"
FT   misc_feature    154747..154747
```



```
FT                   /note="BAM: Bam H1 I/A"
FT   misc_feature    complement(156707..156707)
FT                   /note="polyA signal: AATAAA; 3' end of 2.5kb late (gB) RNA
FT                   and 1.8kb late RNA"
FT   CDS             complement(153702..156746)
FT                   /note="BALF5 DNA polymerase (early), homologous to many DN
A
FT                   polymerases, CMV HFLF2 and RF 28 VZV. 4.5kb early RNA
FT                   apparently encodes BALF5, RNA ends unknown."
FT   promoter        158204..158204
FT                   /note="TATA: TATAAAA"
FT   CDS             complement(156752..159322)
FT                   /note="BALF4 late reading frame 9xNXT/S homologous to HSV1
FT                   glycoprotein B (Pellet et al, 1985), CMV HFLF1 and RF 31
FT                   VZV (gpII)"
FT   promoter        complement(159370..159370)
FT                   /note="TATA: TATTTAA ECL1 late promoter before BALF4, give
s
FT                   2.5kb late RNA"
FT   misc_feature    160966..160966
FT                   /note="polyA signal: AATAAA"
FT   misc_feature    160990..160990
FT                   /note="polyA signal: ATTAAA"
FT   misc_feature    complement(161013..161013)
FT                   /note="polyA signal: AATAAA, presumed end of 3.9kb early
FT                   RNA"
FT   CDS             complement(159312..161678)
FT                   /note="BALF3 reading frame"
FT   misc_feature    163978..166635
FT                   /note="DEL: deletion in Raji"
FT   CDS             complement(161387..164770)
FT                   /note="BALF2 early reading frame, homologous to RF 29 VZV
FT                   and major DNA binding protein HSV. 3.9kb early RNA"
FT   promoter        complement(164814..164814)
FT                   /note="TATA: CATTTAA before BALF2, presumed promoter for
FT                   3.9kb early RNA"
FT   misc_feature    complement(164851..164851)
FT                   /note="polyA signal: AATAAA"
FT   promoter        165442..165442
FT                   /note="TATA: GATAAAA"
FT   promoter        165466..165466
FT                   /note="TATA: TATAAGA EDR1 early promoter before BARF1,
FT                   gives 0.8kb early RNA"
FT   CDS             165504..166169
FT                   /note="BARF1 reading frame"
FT   CDS             complement(164858..165517)
FT                   /note="BALF1 early reading frame, 0.7kb early RNA"
FT   promoter        complement(165713..165713)
FT                   /note="TATA: TATAAAG before BALF1"
FT   misc_feature    166165..166165
FT                   /note="polyA signal: AATAAA 3' end of 0.8kb early RNA. Als
o
FT                   1kb late RNA in this region."
FT   promoter        166469..166475
FT                   /note="TATA: TTATTTT"
FT   mRNA            166498..166916
FT                   /note="Exon 1 of terminal protein RNA"
FT   misc_feature    166561..166563
FT                   /note="Likely initiator met of terminal protein"
FT   misc_feature    166614..166614
```

```
FT                   /note="BAM: Bam H1 A/Nhet"
FT   misc_feature    complement(166946..166946)
FT                   /note="polyA signal: AATAAA These 2 polyA sites are 3' end
FT                   of 0.8kb early, 2.5kb late and 2.5kb latent RNAs"
FT   misc_feature    complement(166950..166950)
FT                   /note="polyA signal: AATAAA"
FT   CDS             complement(167001..167303)
FT                   /note="BNLF2b reading frame"
FT   promoter        167320..167320
FT                   /note="TATA: CATAAAA"
FT   CDS             complement(167307..167486)
FT                   /note="BNLF2a reading frame"
FT   promoter        complement(167525..167525)
FT                   /note="TATA: TATAAAA EDL2 early promoter before BNLF2a,b.
FT                   Gives 0.8kb RNA"
FT   repeat_region   168399..168574
FT                   /note="5 x 33bp repeats"
FT   mRNA            complement(168163..168965)
FT                   /note="BNLF1 coding part of exon c of latent membrane
FT                   protein"
FT   misc_feature    complement(168966..168966)
FT                   /note="ACCEPT: TTTTTTCAACGCAGT possible acceptor"
FT   misc_feature    complement(169041..169041)
FT                   /note="DONOR: TTGGTAAGA donor"
FT   intron          complement(168966..169041)
FT                   /note="intervening sequence in BNLF1"
FT   mRNA            complement(169042..169128)
FT                   /note="BNLF1 exon b of latent membrane protein mRNA"
FT   misc_feature    complement(169129..169129)
FT                   /note="ACCEPT: TCCTCTTTCCCCAGT acceptor"
FT   promoter        complement(169201..169201)
FT                   /note="TATA: TATTACA EDL1A late promoter, gives 2.5kb late
FT                   RNA"
FT   misc_feature    complement(169206..169206)
FT                   /note="DONOR: TGAGTAAGT donor"
FT   intron          complement(169129..169206)
FT                   /note="intervening sequence in BNLF1"
FT   mRNA            complement(169207..169474)
FT                   /note="BNLF1 exon a of latent membrane protein mRNA"
FT   promoter        complement(169546..169546)
FT                   /note="TATA: TACATAAGC EDL1 promoter before BNLF1 gives
FT                   2.5kb latent RNA (LMP)"
FT   repeat_region   170094..170631
FT                   /note="terminal repeat 1 538bp"
FT   repeat_region   170632..171154
FT                   /note="terminal repeat 2 523bp"
FT   repeat_region   171155..171692
FT                   /note="terminal repeat 3 538bp"
FT   repeat_region   171693..172231
FT                   /note="terminal repeat 4 538bp"
XX
SQ   Sequence 172281 BP; 34054 A; 52511 C; 50755 G; 34961 T; 0 other;

Initial Score    =    170  Optimized Score  =    454  Significance = 10.52
Residue Identity =    50%  Matches          =    529  Mismatches   =   418
Gaps             =    105  Conservative Substitutions              =     0

            X       10          20        30        40        50
            CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
            ||| |||  ||    ||| ||  | |    ||    |  ||   ||    || |  |    |
```



```
    CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
107030    107040    107050    107060    107070    107080    107090    107100

    60        70          80          90          100         110         120
    CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
     |    | |||| | | ||  ||| ||||  ||  |    ||| | |   | || ||       |  |||||
    GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
        107110    107120    107130    107140    107150    107160    107170

      130       140       150       160       170       180       190
    CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
    | |   | ||  ||      |     | |  | | ||  || |  ||   ||  |  || || | | | | ||
    CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
          107180    107190    107200       107210    107220    107230

      200       210       220          230       240       250         260
    -TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
     ||  |   |       |   | |    ||| |    |||||| || || |  |  ||||   ||  ||
    GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
107240    107250      107260    107270    107280    107290    107300

          270         280         290       300       310       320
    CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
     |   |||| |||  | |  ||| | ||     ||| | ||||||   |      ||| | || | | |    |
    GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
107310    107320    107330    107340    107350          107360    107370

     330       340       350       360       370         380       390
    CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
      ||        ||  | |||||  |  |||  || |  |||||  ||||||  |   | |||||| |||||  ||
    GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
     107380    107390    107400    107410    107420    107430    107440

      400       410       420       430       440       450         460
    GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
    |||| ||| ||||||   | |||||  |  |  ||  | ||    |||| || || |||    || | ||||
    GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
      107450    107460    107470    107480    107490    107500    107510

          470         480       490       500       510       520
    GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
     | ||  ||  ||||   ||  ||| | |    ||||   ||||    ||| | |  | || | | |  |
    AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
      107520    107530    107540    107550    107560    107570    107580

    530       540       550       560       570       580       590       600
    CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
      ||||   || | ||  ||| |   |  || |    |||||||| ||| | |  || | ||| |    |
    AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
    107590    107600       107610    107620    107630    107640    107650

          610         620         630       640       650       660
    AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
    || |||  ||||  || ||  ||  ||||    ||   || | ||  | | |      |     | | ||
    AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
      107660    107670    107680    107690    107700    107710    107720

     670       680       690       700       710       720       730
    AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
    |||| |   |||    || |  |||  ||  | |  |||  ||   | |  | |  | | |||||    |
```

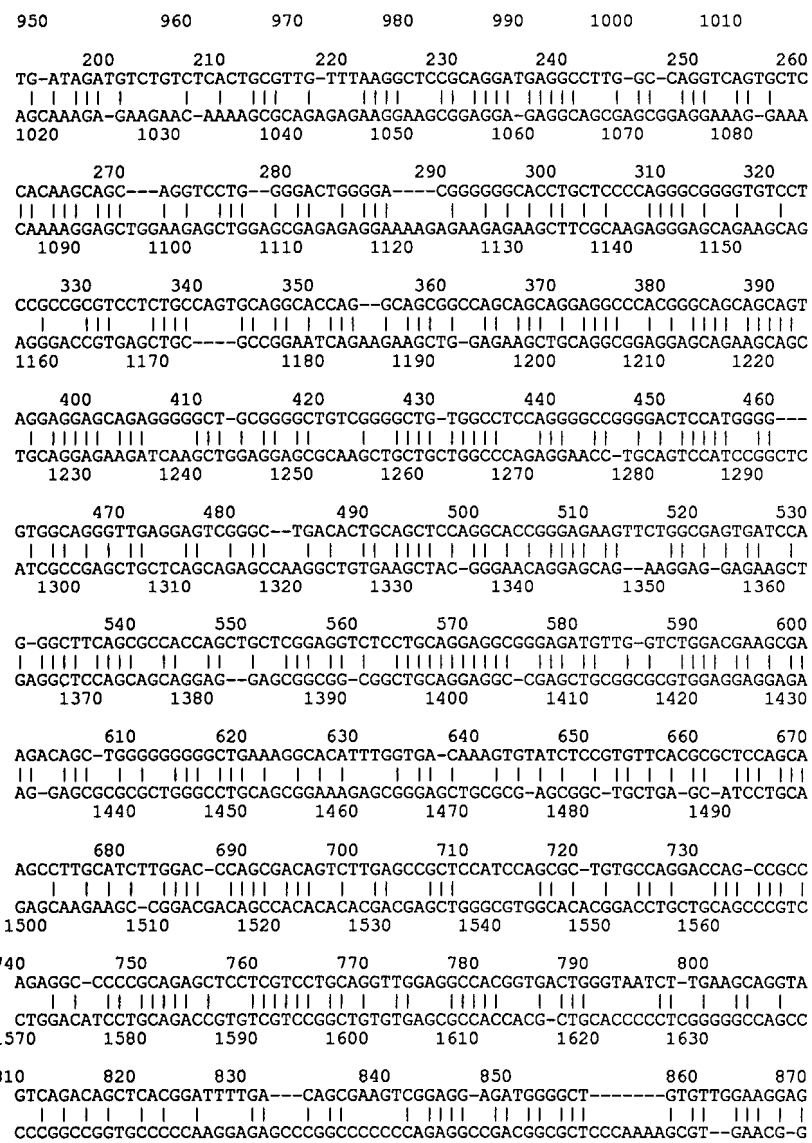
```
    AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
      107730    107740    107750    107760    107770    107780    107790

          740       750       760       770       780           790
    AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA----CTGGGTAATCTT
    ||  ||  ||||      ||| |||    |  |  | | |||  | |||  | ||  |     | ||      |
    AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
    107800    107810    107820    107830       107840    107850    107860

  800       810       820       830       840       850       860
    GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
    || |  || ||   |||    |  || |||     | || | | ||  ||||| |  || || |    || |||
    GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
        107870    107880       107890    107900    107910    107920

  870       880       890       900       910       920       930
    AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
    |||||     ||   | | | ||  ||||    ||||  |  || || | |  ||  |||| | |||   || |
    AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
07930    107940    107950       107960    107970    107980    107990

  940       950       960           970       980       X
    GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
    | | | |   | | ||  |   || | |||    || |  | ||   || |  |
    GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
  108000    108010    108020    108030    108040    108050
```

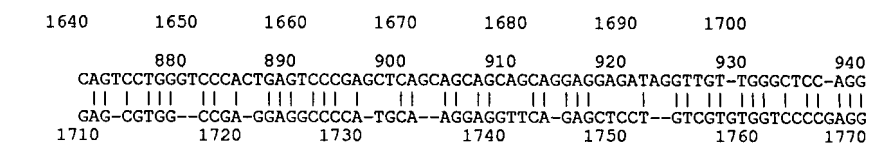
12. US-08-162-407-5' (1-988)
    HUMLSPRO Human lymphocyte surface protein exons 1-5, comple

```
LOCUS       HUMLSPRO     2622 bp ss-mRNA            PRI       25-JAN-1993
DEFINITION  Human lymphocyte surface protein exons 1-5, complete cds.
ACCESSION   M99578
KEYWORDS    lymphocyte surface protein.
SOURCE      Homo sapiens (library: lambda gt11) neonate placenta cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 2622)
  AUTHORS   Voland,J.R., Wyzykowski,R.J., Mark,H. and Dutton,R.W.
  TITLE     Cloning and sequencing of a trophoblast- endothelial- activated
            lymphocyte surface protein: cDNA sequence and genomic structure
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89, 10425-10429 (1992)
  STANDARD  full automatic
COMMENT     NCBI gi: 187241
FEATURES             Location/Qualifiers
     source          1..2622
                     /organism="Homo sapiens"
                     /cell_type="trophoblast"
                     /dev_stage="neonate"
                     /sequenced_mol="cDNA to mRNA"
                     /tissue_type="placenta"
                     /tissue_lib="lambda gt11"
     mRNA            join(1..153,153..933,934..1082,1083..1323,1324..2622)
     exon            1..153
                     /note="codes the 5' UTR; intron between exon 1 and exon 2
                     is 2.7Kb"
                     /number=1
     5'UTR           1..171
```



```
                     /note="84% GC in 5'UTR"
     exon            153..933
                     /note="codes the extracellular domain; intron between exon
                     2 and exon 3 is 0.7Kb"
                     /number=2
     CDS             170..1822
                     /note="550 amino acids MW=61kDa, glycosylated=75 kDa;
                     expressed on endothelium, activated lymphocytes and
                     syncytiotrophoblast, contains leucine zipper and basic
                     region homologous to myc; 721P;  NCBI gi: 187242"
                     /codon_start=1
                     /translation="MAAATIVHDTSEAVELCPAYGLYLKPITKMTISVALPQLKQPGK
                     SISNWEVMERLKGMVQNHQFSTLRISKSTMDFIRFEGEVENKSLVKSFLACLDGKTIK
                     LSGFSDILKVRAAEFKIDFPTRHDWDSFFRDAKDMNETLPGERPDTIHLEGLPCKWFA
                     LKESGSEKPSEDVLVKVFEKFGEIRNVDIPMLDPYREEMTGRNFHTFSFGGHLNFEAY
                     VQYREYMGFIQAMSALRGMKLMYKGEDGKAVACNIKVSFDSTKHLSDASIKKRQLERQ
                     KLQELEQQREEQKRREKEAEERQRAEERKQKELEELERERKREEKLRKREQKQRDREL
                     RRNQKKLEKLQAEEQKQLQEKIKLEERKLLLAQRNLQSIRLIAELLSRAKAVKLREQE
                     QKEEKLRLQQQEERRRLQEAELRRVEEEKERALGLQRKERELRERLLSILQSKKPDDS
                     HTHDELGVAHGPAAARPGHPADRVVRLCERHHAAPPRGPAPGRCPQGEPGPPRGRRRS
                     QKRERERGRGGPMQGGSELLSCGPRGWLSREEVPGRRPLLHS"
     misc_feature    172..232
                     /note="codes for protein leader sequence"
     exon            934..1082
                     /note="codes the putative basic region; intron between
                     exon 3 and exon 4 is 3.5Kb"
                     /number=3
     exon            1083..1323
                     /note="codes the leucine zipper region; the intron between
                     exon 4 and exon 5 is 1.3Kb"
                     /number=4
     misc_feature    1199..1263
                     /note="codes for leucine zipper, homologous to myc"
     exon            1324..2622
                     /note="codes the serine phosphorylation site and the
                     3'UTR"
                     /number=5
BASE COUNT      589 a    792 c    898 g    343 t
ORIGIN

Initial Score    =    167  Optimized Score  =    464  Significance = 10.29
Residue Identity =    51%  Matches          =    546  Mismatches   =   399
Gaps             =    107  Conservative Substitutions               =     0

            X       10        20        30        40          50
            CGGGGGCTTTGTTCAGAGTTGGGTAAATGGGGCCGGTGCTGG--ATTCTGGCGTTGC-CTCC
             ||    | ||         || ||| |||||     | | ||  ||    ||  ||| |
  TTGAACTTCGAGGCCTATGTGCAGTACCGTGAGTACATGGGCTTCATCCAGGCCATGAGCGCCCTGCGCGGG
    810     X 820       830       840       850       860       870

 60        70        80          90        100       110       120
  ATTCTGGGGAGGGA-ATGAGAGGGGCC--CAAGGAAGGGACCGGCCTCTAGCCAACTTCCTCTGTGTCCAGG
  ||       | | | | | | | | || |  ||||| | | || ||  | | | |  || || | ||    |
  ATGAAACTCATGTACAAGGGCGAGGACGGCAAGGCCGTGGCCTGCAAC-ATCAAGGTTTCT-TTGATTCGA
  880       890       900       910       920       930       940

 130       140       150       160       170       180       190
  CTATACATCCTCTGGCTGGTGACTCCCCTCTG-TGTG-CCTCAGTA---TCCTCCCCTGTA--AAATGG-GA
  | | ||| |||  |  || || |||   | | | | | | | |    |     ||  |  || ||| |
  CCAAACA-CCT--GAGTGATGCCTCAATTAAGAAGCGGCAGCTGGAGAGGCAGAAGCTTCAGGAACTGGAGC
```





13. US-08-162-407-5' (1-988)
STMCARA Streptomyces thermotolerans carbomycin-resistance

```
LOCUS       STMCARA      2393 bp ds-DNA             BCT       11-SEP-1992
DEFINITION  Streptomyces thermotolerans carbomycin-resistance protein (carA)
            gene, complete cds.
ACCESSION   M80346
KEYWORDS    carA gene; carbomycin resistance protein.
SOURCE      Streptomyces thermotolerans DNA.
  ORGANISM  Streptomyces thermotolerans
            Prokaryotae; Firmicutes; Streptomycetaceae.
REFERENCE   1  (bases 1 to 2393)
  AUTHORS   Schoner,B.E., Geistlich,M., Rosteck,P., Rao,R.N., Seno,E.,
            Reynolds,P., Cox,K., Burgett,S. and Hershberger,C.L.
  TITLE     Sequence similarity between macrolide resistance determinants and
            ATP-binding transport proteins
  JOURNAL   Gene 115, 93-96 (1992)
  STANDARD  full automatic
COMMENT     NCBI gi: 153197
FEATURES             Location/Qualifiers
     source          1..2393
                     /organism="Streptomyces thermotolerans"
                     /sequenced_mol="DNA"
     CDS             411..2066
                     /gene="carA"
                     /note="NCBI gi: 153198"
                     /codon_start=1
                     /product="carbomycin resistance protein"
                     /translation="MSTAQLALHDITKRYQDHVVLDRIGFTIKPGEKVGVIGDNGSGK
                     STLIKLIAGREQPDNGAVTVVAPGGVGYLAQTLELPLEATVQDAVDLALADLRELEEG
                     MRRTEAELAERPYQTGQDPELAGLLESYAALVDRYQARGGYEADSRVEIALHGLGLPG
                     LERGRRLGTLSGGERSRLALAATLASEPELLLLDEPTNDLDDRAVDWLEEHLRKHKGT
                     VVAVTHDRLFLDRLTTTILEIDSGKVMRYGNGYEGYLAAKAAERQRRLLEYEQWRAEL
                     DRSRDLIASNVARLDAIPRKLPFAVFGAGQFRMRGRGHGAMVRIRNAKERVARLTENP
                     VAPPPEPLTFTAEITTEAAQSRETVAELTGVRVGDRLSVDSLHLGPGERLLVTGPNGA
                     GKTTLLRVLSGELEPDSGSLLVSGRVGHLRQEQTPWRPGMTVLQAFSSGRAGDIDEHT
                     EALLSLGLFSPDDLRQRVQDLSYGQRRRIELARLVTEPVDLLLLDEPTNHLSPALVEE
                     LEEALTGYQGTVVVVTHDRRMRSRFNGAHLTLQDGRVAEFTAA"
BASE COUNT      379 a    823 c    878 g    313 t
ORIGIN
```

```
Initial Score    =    164  Optimized Score  =    448  Significance = 10.07
Residue Identity =     50% Matches          =    540  Mismatches   =   398
Gaps             =    131  Conservative Substitutions              =     0
```

```
           X       10        20        30        40        50
           CGGGGGCTTTGTTCAGAGTTGGGTA--AATGGGGC-CGGTGCTG---GA--TTCTGGCGTTG
           ||  ||    | || |    |   |  || | | ||  ||||   ||    | ||    |
  GGCCATGGTGCGATGGTGCGGATCCGCAACGCCAAGGAACGCGTCGCGCGGCTGACCGAAAACCCGGTCGCG
  1350      1360      1370      1380      1390      1400      1410

      60        70              80        90       100       110       120
  CCTCCATTCTGGGGAGGGAA---TGA--GAGGGGCCCAAGGAAGGGACCGGCCTCTAGCCAACTTCCTCTGT
  || ||   | | |  |   |   | |  | || |  ||     | | |||  |   || |        | ||
  CCGCC-GCCCGAGCCGCTCACCTTCACCGCGGAGATCACCACCGAGGCCGCGC---AGTCCCGGGAGACGGT
1420      1430      1440      1450      1460      1470      1480

       130       140       150       160       170       180
  GTCCAGGCTATACATCCTCTGGCT-GGTGACTCCCCTCTGTGT-GCCTCAGT-ATCCTC---CCCTGTAAAA
  | ||   ||   |  | || || | || ||| |  ||| | || | |||  |   ||||   ||| || ||
  GGCCGAACTCACCGGCGTCCGGGTCGGCGAC-CGGCTCAGCGTCGACTCCCTGCACCTCGGGCCCGGTGAA-
  1490      1500      1510      1520      1530      1540      1550

  190       200       210       220       230       240       250
  TGGGATGATAGATGTCTGTCTCACTGCGTTGTTTAAGGCTCCGCAGGATGAGGCCTTGGCCAGGTCAGTGCT
    || || | |    | | | ||  | |  |   ||| |  ||| |     ||   |  || ||    ||
  -CGGCTGCTGGTCACCGGCCCCAACGGGGCGGGCAAGACGACGCTGCTGCGGGTGCTCTCCGGGGAGCTGGA
   1560      1570      1580      1590      1600      1610      1620

260       270       280       290       300       310       320
  CCACAAGCAGCAGGT-CCTG-GGGACTGGGGACGG--GGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCGC
   | | | |||| | | ||||  ||  | ||| |||  ||| ||||||       |||  ||       | ||
  ACCCGA-CAGCGGATCCCTGCTGGTGTCGGGCCGGGTGGGACACCTG------AGGCAGGAACAGACACCAT
 1630      1640      1650      1660      1670              1680      1690

  330       340       350       360       370       380       390
  CGCGTCC---TCTGCCAGTG---CAGGCACCAGGCAGCGGCCAGCAGCAGGAGGC--CCACGGGCAGCAGCA
   ||| ||   | || | |||   |||||     |||||||  | |  || || | |  | ||| ||| || |
  GGCGGCCGGGTATGACGGTGCTTCAGGCGTTCAGCAGCGGACGG--GCCGGTGACATCGACGAGCA-CACC-
      1700      1710      1720      1730      1740      1750      1760

       400       410       420       430       440       450
  GTAGGAGGAGCAGAGGGGGCTGCGGGGCTG-TCGGGGCTGTGG-CCT---CCAG-GGGCCGGGGACTCCATG
      ||||  || |       ||  || |||||  || |  | |  | |||   |||| | |    |||| |  |
  ----GAGGCGCTGCTCTCCCT-CG-GGCTGTTCAGCCCGGACGACCTCCGCCAGCGTGTGCAGGAC-CTGTC
             1770      1780      1790      1800      1810      1820

  460       470       480       490       500       510       520
  GGGGTGGCAGGGTTGAGGAGTCGGGCTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTG--GC-GAGTGA
  |    ||||| |  |  |  ||| |||  | | || | |   ||  ||||  ||  | |||  || |   ||
  GTACGGGCAGCGGCGCCGCATCGAGCTTGCCCGGCTGGTGACGGAGCCGGTCGACCTGCTGCTGCTGGACGA
   1830      1840      1850      1860      1870      1880      1890

   530       540       550       560       570       580       590
  TCCAGGGCTTCAGCGC-CACCAGCTGCTCG--GAGGTCTCCTGCAGGAGGCG-GGAGATGTTGGTCTGGACG
   ||       || | | | || || |||||  ||||     ||| ||||||||  ||    |||  || || |
  ACCCACCAACCACCTCTCGCCCGC-GCTCGTCGAGG--AACTGGAGGAGGCGCTGACCGGTT-ATCAGGGC-
 1900      1910      1920      1930      1940      1950      1960

     600       610       620       630       640       650       660
  AAGCGAAGACAGCTGGGGGGGGGCTGAAAGGCACATTTGGTGACAAAGTGTATCTCCG-TGTTCACGCGCTC
  | | |  | |   | |       | ||  ||| ||| |||  |    |  | |  ||    ||   ||||
  ACG-GTCGTC--GTCGTCACCCAC-GACCGGCGCATGCGGT--CCCGCTTCAACGGCGCCCATCTGACGCTG
     1970      1980      1990      2000      2010      2020      2030
```



```
   670       680       690       700       710       720       730
  CAGCAAGCCTTGCATCTTGGA-CCCAGCGACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CCA
  ||| | |    || ||   ||  || || | | | |||  ||| | |   ||||  || | ||||    |
  CAGGACG-GGCGCGTCGCCGAGTTCACCG-CCGCC-TGA-TCGC-CGA---AGCG--GTACGAGGAGGGACG
        2040      2050      2060      2070              2080      2090

      740               750           760       770           780       790
  GCCGCCAGAGGCCC------CCGCAGAGCTC----CTCGTCCTGCAG----GTTGGAGGCCACGGTGACTGG
  |  ||||| |||||       |||||  | ||      | || ||||     |  || |||||  |   | ||
  GGTGCCAGGGGCCCGTCCCTCCGCA-CGGTCAGAGACGGACCGGCAGACTCTTCAGACGCCAC-GCACCCGG
     2100      2110      2120      2130      2140      2150      2160

       800       810       820       830       840       850
  GTAATCTTGAAGCAGGTAGTCAGACAGC-TCACG--GATTTTGA--CA-GCGAAGTCGGAGGAGATGGGGCT
  | |    ||  |       | | ||| ||  ||||  ||    ||  |  | ||||  |  | ||| || |
  GAAGGGCTGCCGTTCCAGGGCGGACGGCGGCACGGCGAGGGCGAGGCTGGGGAAGCGGTCGAAGAGGGTGTC
     2170      2180      2190      2200      2210      2220      2230

  860       870       880       890       900       910       920
  GTGTTGGAAGGAGCAGTCCTGG-GTCCCACTGAGTC-CCGAGCTCAGCAGCAGCAGGAGGAGATAG--G
  | |   ||  | |  ||||||  || |||  | | | || ||  ||| ||     || | ||  || ||  |
  GAGGGCGATCGCGGTCTCCTGGTGTGCCAGGGCGGCGCCCAG-GCAGTAGTGCGGGCCGAAGCCGA-AGCCG
    2240      2250      2260      2270      2280      2290      2300

    930       940       950       960       970       980       X
  TTGTTGGGCTCCAG-GCTG-GCGCCAG---CACTGTCATTTCG-GCCG--GGGGCCCCGGAATTCCGGCCG
    ||  | |||  | | || |||   |   |   || || |||  |||   |||  | ||    |||| |
  ACGTGTGCCTCGTGCGGTGCGCGGTGGTTTCCGGGTGATGTCGAACCGCTCGGGGTCGGGGGAACCGGTCCG
    2310      2320      2330      2340      2350      2360      2370    X

  GATCGTGTC
  2380
```

14. US-08-162-407-5' (1-988)
    GORINVOLUB Gorilla gorilla involucrin gene medium allele, com

```
LOCUS       GORINVOLUB   1818 bp ds-DNA             PRI       08-SEP-1994
DEFINITION  Gorilla gorilla involucrin gene medium allele, complete cds.
ACCESSION   M23604 J04499
KEYWORDS    epidermal protein; involucrin.
SOURCE      Gorilla gorilla (individual_isolate Gorilla K) (library: vagina)
            DNA.
  ORGANISM  Gorilla gorilla
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Pongidae.
REFERENCE   1  (bases 1 to 1818)
  AUTHORS   Teumer,J. and Green,H.
  TITLE     Divergent evolution of part of the involucrin gene in the
            hominoids: Unique intragenic duplications in the gorilla and human
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86, 1283-1286 (1989)
  STANDARD  full automatic
COMMENT     Draft entry and computer readable copy of sequence [1] kindly
            provided by J.K. Teumer (03-31-89).

            NCBI gi: 340978
FEATURES             Location/Qualifiers
     source          1..1818
                     /organism="Gorilla gorilla"
                     /isolate="Gorilla K"
```

```
                     /cell_type="keratinocyte"
                     /sequenced_mol="DNA"
                     /clone="pIGorH6.1"
                     /tissue_lib="vagina"
     CDS             1..1818
                     /note="NCBI gi: 536828"
                     /codon_start=1
                     /product="involucrin"
                     /translation="MSLQHTLPVTLSPALSQELLKTVPPPVNTQQEQMKLPTPLPPPC
                     QKVPVKLPVEVPSKQEEKHMTAVKGLPEQECEQQQQEPQEQEPQQQHWEQHEEHQKAE
                     NPEQQLKQEKAQRDQQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELPEQRE
                     GQLKHLEQQKGQLELPEQQEGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLELPEQ
                     QEGQLELPEQQEGQLKHLEQQEGQLELPEQQEGQLELPEQQEGQLKHLEQQEGQLKHL
                     ENQEGQLEVPEEQVGQLKYLEQQEGQLKHLDQQEKQPELPEQQVGQLKHLEQQEGQPK
                     HLEQQEGQLEHLEEQEGQLKHLEQQEGQLEHLEHQEGQLGLPEQQVRQLKQLEKEEGQ
                     PKHLEEEEGQLKHLVQQEGHVEHLVQQEGQLEHLVQQEGQLEQQEGQVEHLEQQVGQL
                     KHLEEQEGQLKHLEQQQGQLGVPEQQVGQPNNLEQEEKQLELPEQQEGQLKHLKKQEA
                     QLELPEQQVGQPKHLEQQEKQLEHPEQQDGQLKHLEQQEGQLKDLEQQKGQLEQPVFA
                     PAPGQVQDIQPALPTKGELLLPIEHQQQKQEVQWPPKHK"
 BASE COUNT      531 a    454 c    657 g    176 t
 ORIGIN

Initial Score    =    160  Optimized Score  =    462  Significance =  9.76
Residue Identity =    51%  Matches          =    555  Mismatches   =   379
Gaps             =    134  Conservative Substitutions               =     0

             X          10         20         30         40         50
             CGGGG--GCT-TTGTTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTGCCTCC
              ||||  |||   | ||  ||    | |   |||||    ||||||  ||  |   | | |
   GAGCAGCAGGAGGGGCAGCTGGAGCTCCCAGAGCAGCAGGAGGGGC---AGCTGGA-GCTCCCAGAG-CAGC
  710        720        730        740        750          760        770

  60         70         80         90        100        110
   ATTCTGGGGAG-GGAATGA-GAGGGGCCCAAGGAAGGG-ACC---GGCCTCTAGCCAAC-------TTCCTC
   |    ||  ||  |||  |   || ||   |||| ||| | |    || ||| |  |||         |  |
   AGGAGGGACAGCTGAAGCACCTGGAGCAGCAGGAGGGGCAGCTGAAGCATCTGGAGAACCAGGAGGGGCAGC
     780        790        800        810        820        830        840

  120          130        140        150          160        170        180
   TG--TGTCCAGGCTATACATCCTCTGGCTGGTGACTCCCCT----CTG--TGTGCCTCAGTATCCTCCCCTG
   ||   ||||  |    ||   |  ||| | |||    |||    | |   | |  |||       | ||||
   TGGAGGTCCCAGAGGAGCA-GGTGGGGCAGCTGAAGTACCTGGAACAGCAGGAGGGGCAGCTGAAGCACCTG
   850        860        870        880        890        900        910

        190        200        210        220        230        240        250
   TAAAATGGGATGATAGATGTCTG-TCTCACTGCGTTGTTTAAGGCTCCGCAGGATGAGGC-CTTGG-CCAG-
    |   |   || ||  | | |  ||| | | |   |||    ||| | ||| || | ||| ||  |||
   GATCAGCAGGAGA-AGCAGCCGGAGCTCCCAGAGCAG---CAGGTGGGGCA-GCTGAAGCACCTGGAGCAGC
  920        930        940        950          960        970        980

           260        270        280        290        300        310
   -GTCAGTGCTCCACAAGCAGCAGGTCCTGGGGACTGGGGACGGGGGGCACCTGCTCCCCAGG-GC-GGGGTG
    |   | ||  | ||||| | ||  | |  |   ||| |   || ||||||||       ||| || || | |
   AGGAGGGGCAGCCTAAGCATCTGGAGCAGCAGGAGGGGCAACTGGAGCACCTG-----GAGGAGCAGGAGGG
     990       1000       1010       1020       1030            1040       1050

  320        330        340        350        360        370          380
    TCCTCCGCCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCA
     | | |  ||  |||     ||  |||||      ||||| |  |||| | ||||  ||    | ||||||||
    GCAGCTGAAGC-ACCTGGAGCA--GCAGG--AGGGGCAGCTG-GAGCACCTGGAGCACCAGGAAGGGCAGCT
```



```
       1060       1070          1080       1090       1100       1110

  390        400        410        420        430        440        450
   GCAGTAGGAGGAGCAGAGGGGGCTGCGGGGCTGTCGGGGCT-GTG-GCCTCCAGGGGCCGGGGACTC-CATG
   |  |       ||||||  || || ||   ||||  |  ||| | |       ||||||  |   |  | | ||
   GGGGCTCCCAGAGCAGCAGGTGCGGC--AGCTGAAGCAGCTAGAGAAGGAGGAGGGGCAGCCAAAGCACCTG
 1120       1130       1140       1150       1160       1170       1180

   460        470        480        490        500        510        520
   GGGGTGGCAGGGTTGAGGAGTCG-GGCTGACACTGCAGCTCCA--GGCACCGGGAGAAGTTCTGGCGAGTGA
   | || || || |  | | ||| |  ||      ||||||   |  |||||  |||| |   |||| |    |
   GAGGAGG-AG-GAGGGGCAGTTGAAGCACCTGGTGCAGCAGGAGGGGCACGTGGAGCA--CCTGGTGCAGCA
 1190       1200       1210       1220       1230       1240       1250

   530        540        550          560        570        580        590
   TCCAGGGCTTCAGCGCCACCAGCTGC--TCGGA-GGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGA
       ||||  | |   || | | |||    ||| || |  ||| || | || ||||  |  ||  ||||  |
   GGAGGGGCAGCTGGAGCATCTGGTGCAGCAGGAGGGGCAGCTGGAGCA-GCAGGAGGGGCAGG--TGGAGCA
   1260       1270       1280       1290       1300       1310       1320

     600        610        620        630        640        650        660
   AGCGAAGACAGCTGG-GGGGGGGCTGAAAGGCACAT-TTGGTG-ACAAAGTGTATCTCCGTGTTCACGCGCT
      | || |||| || ||||  ||||||  |||| |   || | |  | | | | ||      ||  | ||
   CCTGGAG-CAGCAGGTGGGGCAGCTGAA--GCACCTAGAGGAGCAGGAGGGGCAACTGAAGCATCTGGAGCA
    1330       1340       1350       1360       1370       1380       1390

     670        680        690        700        710        720          730
   CCAGCAAGCCTTGCA--TCTTGGACCCAGCGACAGTCTTGAGCCGC-TCCATCCAGCGCTGTGC--CAGGAC
   |||||  |    |||  |   || |||||  | |||  |  |  ||  |||  || |   | ||  ||||
   GCAGCAGG---GGCAGTTGGGGGTCCCAGAG-CAG-CAGGTGGGGCAGCCAAACAACCTGGAGCAGGAGGAG
     1400       1410       1420       1430       1440       1450       1460

         740        750          760        770        780
   CAGCCGCCAGAGGCCCC--CGCA----GAGCTCCTCGT------CCTGCAG--GTTGGAGGCCACGGTGACT
    ||| ||  |||  |||   |||     |||  ||   |      |||| ||  |  ||||| ||| |   ||
   AAGCAGCTGGAGCTCCCAGAGCAGCAAGAGGGCCAGCTGAAGCACCTGAAGAAGCAGGAGG-CACAG---CT
       1470       1480       1490       1500       1510       1520

  790        800        810        820        830        840        850
   GGGTAATCTTGA--AGCAGGTAGTCAGACAGCTCACGGATTTTGA-CAGCGAAGTCGGAGGAGATGGGGCTG
   ||     |  ||  |||||||||   ||||||  | | |  | || ||||  ||     || || | | ||
   GGAGCTCCCAGAGCAGCAGGTAG---GACAGCCAAAGCACCTGGAGCAGC--AGGAAAAGCAGCTAGAGC--
 1530       1540       1550       1560       1570       1580       1590

  860        870        880        890        900        910        920
   TGTTGGAAGGAGCAGTCCTGGGTCCCACTGAGTCCCGAGC-TCAGCAGCAGCAGCAGGAGGAGATAG-GTTG
      || | |||||  | |    | |      ||  ||||  ||| ||  |||||  | ||||  | | |  |
   ACCCAGA-GCAGCAGGACGGACAACTAAAACATCTGGAGCAGCAGGAGGGGCAGCTGAAGGACCTGGAGCAG
     1600       1610       1620       1630       1640       1650       1660

  930        940        950        960              970        980        X
   TTG-GGCTCCAGGCTGGCGCCAGCACTGTCATTTC-----GGCCGGG---GGCCCCGGA-ATTCCGGCCG
    |  |   || ||||| | |||| ||||  || |         || ||   ||  ||  || ||||   | |
   CAGAAGGGGCA-GCTGGAG-CAGC-CTGTGTTTGCCCCAGCTCCAGGCCAGGTCCAAGACATTCAACCAGCC
   1670       1680       1690       1700       1710       1720       1730 X

   CTGCCCAC
     1740
```

15. US-08-162-407-5' (1-988)
GORINVOLUA Gorilla gorilla involucrin gene large allele, comp

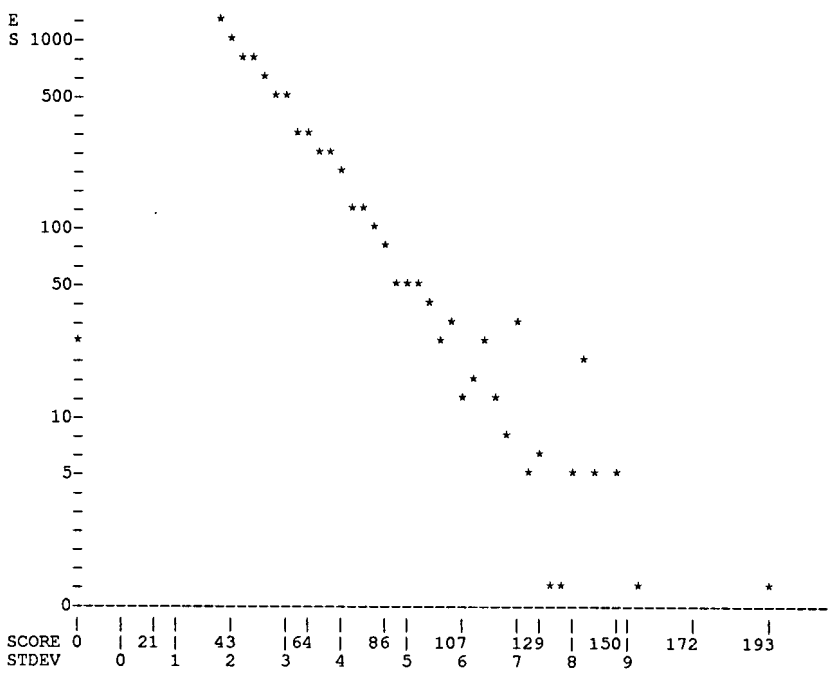
```
LOCUS       GORINVOLUA   1908 bp ds-DNA             PRI       07-SEP-1994
DEFINITION  Gorilla gorilla involucrin gene large allele, complete cds.
ACCESSION   M23603 J04499
KEYWORDS    epidermal protein; involucrin.
SOURCE      Gorilla gorilla (individual_isolate Gorilla F) DNA.
  ORGANISM  Gorilla gorilla
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Pongidae.
REFERENCE   1  (bases 1 to 1908)
  AUTHORS   Teumer,J. and Green,H.
  TITLE     Divergent evolution of part of the involucrin gene in the
            hominoids: Unique intragenic duplications in the gorilla and human
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86, 1283-1286 (1989)
  STANDARD  full automatic
COMMENT     Draft entry and computer readable copy of sequence [1] kindly
            provided by J.K. Teumer (03-31-89).

            NCBI gi: 340977
FEATURES             Location/Qualifiers
     source          1..1908
                     /organism="Gorilla gorilla"
                     /isolate="Gorilla F"
                     /cell_type="fibroblast"
                     /sequenced_mol="DNA"
                     /clone="pIGorH6.2"
                     /note="(vector pGEM3)"
     CDS             1..1908
                     /note="NCBI gi: 536827"
                     /codon_start=1
                     /product="involucrin"
                     /translation="MSLQHTLPVTLSPALSQELLKTVPPPVNTQQEQMKPPTPLPPPC
                     QKVPVKLPVEVPSKQEEKHMTAVKGLPEQECEQQQQEPQEQEPQQQHWEQHEEHQKAE
                     NPEQQLKQEKAQRDQQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELPEQRE
                     GELKHLEQQKGQLELPEQQEGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLELPEQ
                     QEGQLELPEQQEGQLKHLEQQEGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLELP
                     EQQEGQLELPEQQEGQLKHLEQQEGQLKHLENQEGQLEVPEEQVGQLNYLEQQEGQLK
                     HLDQQEKQPELPEQQVGQLKHLEKQEGQPKHLEQQEGQLEHLEEQEGQLKHLEQQEGQ
                     LEHLEHQEGQLGLPEQQVRQLKQLEKEEGQPKHLEEEEGQLKHLVQQEGQLEHLVQQE
                     GQLEHLVQQEGQLEQQEGQVEHLEQQVGQLKHLEEQEGQLKHLEQQQGQLGVPEQQVG
                     QPNNLEQEEKQLELPEQQEGQLKHLEKQEAQLELPEQQVGQPKHLEQQEKQLEHPEQQ
                     DGQLKHLEQQEGQLKDLEQQKGQLEQPVFAPAPGQVQDIQPALPTKGELLLPVEHQQQ
                     KQEVQWPPKHK"
BASE COUNT      554 a    476 c    696 g    182 t
ORIGIN
```

```
Initial Score    =    160  Optimized Score  =    462  Significance =  9.76
Residue Identity =    52%  Matches          =    556  Mismatches   =   377
Gaps             =    136  Conservative Substitutions               =     0

            X          10         20         30         40         50
            CGGGG--GCT-TTGTTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTGCCTCC
             ||||  |||   | ||  ||    | |   |||||    ||||||  ||  |   | |  |
   GAGCAGCAGGAGGGGCAGCTGGAGCTCCCAGAGCAGCAGGAGGGGC---AGCTGGA-GCTCCCAGAG-CAGC
  800        810        820        830        840          850        860

   60         70          80         90          100        110
    ATTCTGGGGAG-GGAATGA-GAGGGGCCCAAGGAAGGG-ACC---GGCCTCTAGCCAAC-------TTCCTC
```



```
     |    ||  ||  |||  |   ||  ||   |||| ||| | |    || ||| |  |||          | |
    AGGAGGGACAGCTGAAGCACCTGGAGCAGCAGGAGGGGCAGCTGAAGCATCTGGAGAACCAGGAGGGGCAGC
      870        880        890        900        910        920        930

   120         130        140        150          160         170        180
    TG--TGTCCAGGCTATACATCCTCTGGCTGGTGACTCCCCT----CTG--TGTGCCTCAGTATCCTCCCCTG
    ||   ||||  |      ||   |  ||| | ||| |  |||     | |   | |  |||       | ||||
    TGGAGGTCCCAGAGGAGCA-GGTGGGGCAGCTGAATTACCTGGAACAGCAGGAGGGGCAGCTGAAGCACCTG
    940        950        960        970        980        990       1000

         190        200        210        220        230        240
    TAAAATGGGATGATAGATGTCTG-TCTCACTGCGTTGTTTAAGGCTCCGCAGGATGAGGC-CTTGG----CC
     |   |    |  || ||  | | |  ||| | | |  |    |||    ||| | ||| || | |||     |
    GATCAGCAGGAGA-AGCAGCCGGAGCTCCCAGAGCAG---CAGGTGGGGCA-GCTGAAGCACCTGGAGAAGC
1010       1020       1030       1040         1050       1060       1070

    250        260        270        280        290        300        310
    AGGTCAGTGCTCCACAAGCAGCAGGTCCTGGGGACTGGGGACGGGGGGCACCTGCTCCCCAGG-GC-GGGGT
    |||   | ||  |  |||||  | ||  | |  |   ||| |   || |||||||       ||| || || |
    AGG-AGGGGCAGCCTAAGCATCTGGAGCAGCAGGAGGGGCAACTGGAGCACCTG-----GAGGAGCAGGAGG
      1080       1090       1100       1110       1120           1130       1140

    320        330        340        350        360        370          380
    GTCCTCCGCCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGC
    | |  | |  ||  |||     ||  |||||      |||||| |  |||| | |||| ||   | |||||||
    GGCAGCTGAAGC-ACCTGGAGCA--GCAGG--AGGGGCAGCTG-GAGCACCTGGAGCACCAGGAAGGGCAGC
         1150       1160         1170       1180       1190       1200

    390        400        410        420        430        440        450
    AGCAGTAGGAGGAGCAGAGGGGGCTGCGGGGCTGTCGGGGCT-GTG-GCCTCCAGGGGCCGGGGACTC-CAT
     |  |      ||||||  || || ||   ||||  |  |||  | |          ||||||  |   |  | | |
    TGGGGCTCCCAGAGCAGCAGGTGCGGC--AGCTGAAGCAGCTAGAGAAGGAGGAGGGGCAGCCAAAGCACCT
1210       1220       1230        1240       1250       1260       1270

     460        470        480        490         500        510        520
    GGGGGTGGCAGGGTTGAGGAGTCG-GGCTGACACTGCAGCTCCA--GGCACCGGGAGAAGTTCTGGCGAGTG
    || || || || |  | | ||| |  ||       ||||||   |  |||| | |||| |   |||| |
    GGAGGAGG-AG-GAGGGGCAGTTGAAGCACCTGGTGCAGCAGGAGGGGCAGCTGGAGCA--CCTGGTGCAGC
1280       1290       1300       1310       1320       1330       1340

      530        540        550         560        570        580        590
    ATCCAGGGCTTCAGCGCCACCAGCTGC--TCGGA-GGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACG
    |    ||||  | |   || | | |||    ||| || |  ||| || | || ||||  |  ||  ||||
    AGGAGGGGCAGCTGGAGCATCTGGTGCAGCAGGAGGGGCAGCTGGAGCA-GCAGGAGGGGCAGG--TGGAGC
     1350       1360       1370       1380       1390       1400       1410

       600        610        620        630        640        650        660
    AAGCGAAGACAGCTGG-GGGGGGGCTGAAAGGCACAT-TTGGTG-ACAAAGTGTATCTCCGTGTTCACGCGC
    |   | || |||| || ||||  ||||||  ||||  |   || | |  | | | | ||       ||  | ||
    ACCTGGAG-CAGCAGGTGGGGCAGCTGAA--GCACCTAGAGGAGCAGGAGGGGCAACTGAAGCATCTGGAGC
      1420       1430       1440       1450       1460       1470       1480

       670        680        690        700        710        720        730
    TCCAGCAAGCCTTGCA--TCTTGGACCCAGCGACAGTCTTGAGCCGC-TCCATCCAGCGCTGTGC--CAGGA
      ||||| |    |||  |   || ||||| | ||| |  | |  ||  |||  || |   | ||   ||||
    AGCAGCAGG---GGCAGTTGGGGGTCCCAGAG-CAG-CAGGTGGGGCAGCCAAACAACCTGGAGCAGGAGGA
      1490       1500       1510       1520       1530       1540       1550

          740        750        760        770        780
    CCAGCCGCCAGAGGCCCC--CGCA----GAGCTCCTCGT------CCTGCAG--GTTGGAGGCCACGGTGAC
```

```
      ||| ||  |||  |||   |||    |||  ||   |      |||| ||  |  ||||| ||| |   |
    GAAGCAGCTGGAGCTCCCAGAGCAGCAAGAGGGCCAGCTGAAGCACCTGGAGAAGCAGGAGG-CACAG---C
          1560       1570       1580       1590       1600       1610

   790        800         810        820        830        840        850
    TGGGTAATCTTGA--AGCAGGTAGTCAGACAGCTCACGGATTTTGA-CAGCGAAGTCGGAGGAGATGGGGCT
    |||      |  ||  |||||||||   ||||||  | | |  | || ||||  ||    || || | | ||
    TGGAGCTCCCAGAGCAGCAGGTAG---GACAGCCAAAGCACCTGGAGCAGC--AGGAAAAGCAGCTAGAGC-
  1620       1630       1640          1650       1660       1670       1680

    860        870        880        890        900        910        920
    GTGTTGGAAGGAGCAGTCCTGGGTCCCACTGAGTCCCGAGC-TCAGCAGCAGCAGCAGGAGGAGATAG-GTT
         || | |||||  | |    | |     ||  ||||  ||| ||  |||||  | ||||  | | |
    -ACCCAGA-GCAGCAGGACGGACAACTAAAACATCTGGAGCAGCAGGAGGGGCAGCTGAAGGACCTGGAGCA
       1690       1700       1710       1720       1730       1740       1750

    930        940        950        960               970        980        X
    GTTG-GGCTCCAGGCTGGCGCCAGCACTGTCATTTC-----GGCCGGG---GGCCCCGGA-ATTCCGGCCG
    |  |  |   || ||||| | |||| ||||  || |       || ||    || ||  || ||||   | |
    GCAGAAGGGGCA-GCTGGAG-CAGC-CTGTGTTTGCCCCAGCTCCAGGCCAGGTCCAAGACATTCAACCAGC
      1760       1770        1780       1790       1800       1810       1820 X

    CCTGCCCAC
       1830
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq5nsq.res made by  on Wed 5 Apr 95 20:58:10-PDT.


Query sequence being compared:US-08-162-407-5 (1-988)
Number of sequences searched:                57621
Number of scores above cutoff:                4402

     Results of the initial comparison of US-08-162-407-5 (1-988) with:
  Data bank : N-GeneSeq 17, all entries

100000-
      -
N     -
U50000-
M     -
B     -
E     -
R     -
      -
O     - **
F10000-
      -*
S     -   *
E 5000-
Q     -
U     -    *
E     -     ****
N     -         **
C     -           *
```

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	19	14	15.41

Times:	CPU	Total Elapsed
	00:04:04.07	00:04:05.00

Number of residues:	24347505
Number of sequences searched:	57621
Number of scores above cutoff:	4402

Cut-off raised to 11.
Cut-off raised to 18.
Cut-off raised to 26.
Cut-off raised to 32.
Cut-off raised to 38.
Cut-off raised to 42.
Cut-off raised to 47.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

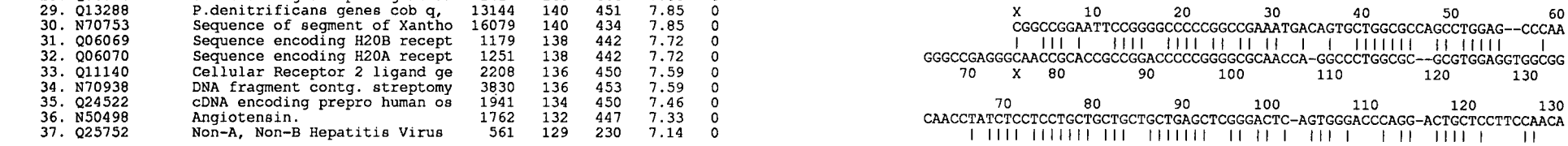
	Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
		**** 11 standard deviations above mean ****					
1.	Q41760	Oxytocin receptor coding sequ	1167	193	455	11.29	0
		**** 8 standard deviations above mean ****					
2.	Q14645	Clone associated with biochem	2702	155	453	8.83	0
3.	Q05536	Sequence enoding purified sol	1443	150	453	8.50	0
4.	Q21171	Clone encoding Intercellular	1846	149	454	8.44	0
5.	N90611	Intercellular adhesion molecu	1846	149	454	8.44	0
6.	Q04049	Sequence of clone HL-60 encod	3024	149	454	8.44	0
7.	N80995	HL-cDNA clone encoding human	3113	149	451	8.44	0
8.	Q06583	Sequence encoding bovine meta	1045	143	326	8.05	0
9.	Q55132	Sequence encoding enzyme whic	5535	143	449	8.05	0
		**** 7 standard deviations above mean ****					
10.	Q38735	Human OP-2.	1723	142	393	7.98	0
11.	Q01758	cDNA sequence of rat C kinase	2909	142	428	7.98	0
12.	Q55131	Sequence encoding enzymes whi	5533	142	450	7.98	0
13.	Q55342	tICAM(452) cysteine mutant.	1359	141	453	7.92	0
14.	Q55340	tICAM(453)IgG immunoadhesin.	2043	141	454	7.92	0
15.	Q57016	PKC delta.	2891	141	431	7.92	0
16.	Q57397	Human ICAM-1.	3024	141	454	7.92	0
17.	Q45164	Human OP-2.	1723	140	394	7.85	0
18.	Q65393	Osteogenic protein hOP2-PP.	1723	140	392	7.85	0
19.	Q45118	Human OP-2.	1723	140	394	7.85	0
20.	Q67313	Human OP-2.	1723	140	394	7.85	0
21.	Q57917	Human osteogenic protein hOP2	1723	140	393	7.85	0
22.	Q56233	hOP2 cDNA.	1723	140	393	7.85	0
23.	Q56200	hOP2 cDNA.	1723	140	393	7.85	0
24.	Q53155	Sequence encoding human osteo	1723	140	393	7.85	0
25.	Q38859	Morphogen hOP2 coding sequenc	1723	140	392	7.85	0
26.	Q38946	Human osteogenic protein 2 (h	1723	140	393	7.85	0
27.	Q28737	Human osteogenic protein hOP2	1723	140	393	7.85	0
28.	Q56106	DNA encoding a lipase gene an	2417	140	464	7.85	0
29.	Q13288	P.denitrificans genes cob q,	13144	140	451	7.85	0
30.	N70753	Sequence of segment of Xantho	16079	140	434	7.85	0
31.	Q06069	Sequence encoding H20B recept	1179	138	442	7.72	0
32.	Q06070	Sequence encoding H20A recept	1251	138	442	7.72	0
33.	Q11140	Cellular Receptor 2 ligand ge	2208	136	450	7.59	0
34.	N70938	DNA fragment contg. streptomy	3830	136	453	7.59	0
35.	Q24522	cDNA encoding prepro human os	1941	134	450	7.46	0
36.	N50498	Angiotensin.	1762	132	447	7.33	0
37.	Q25752	Non-A, Non-B Hepatitis Virus	561	129	230	7.14	0

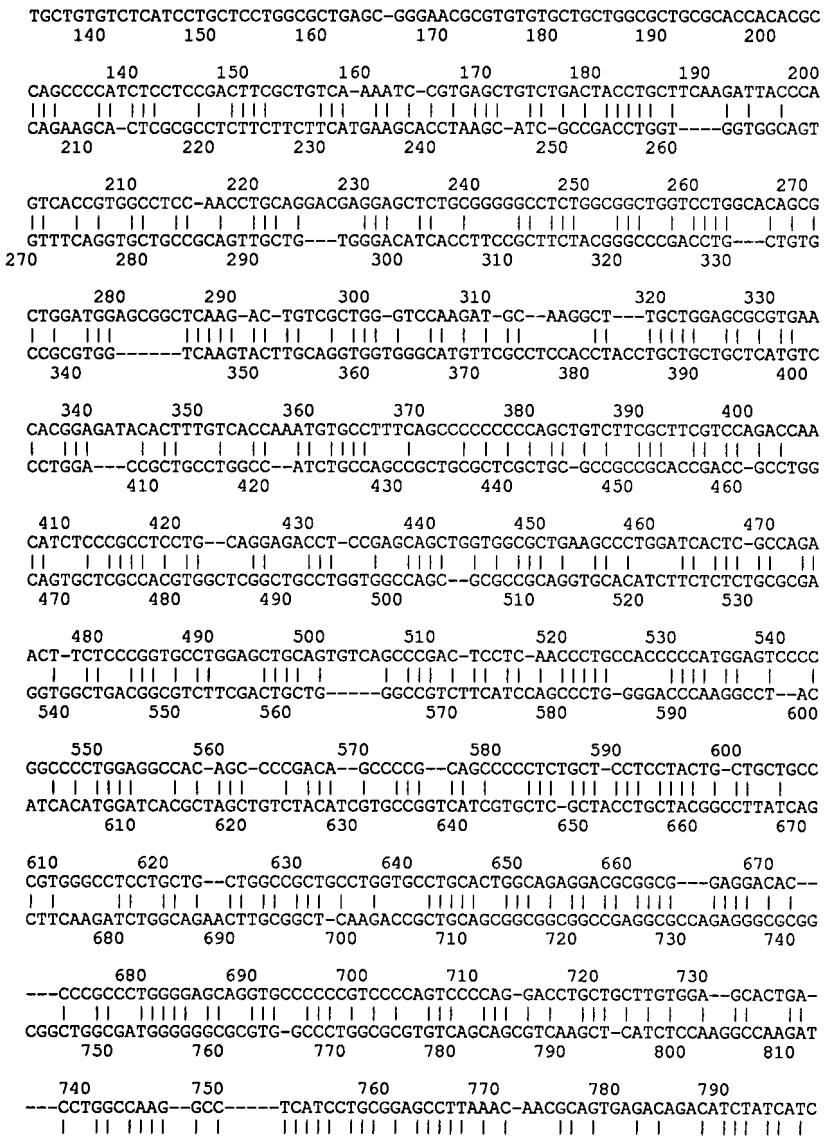
38.	Q47954	Calf IAP DNA.	5399	129	445	7.14	0
39.	Q47065	tie truncated receptor kinase	3713	128	455	7.07	0
40.	Q47064	tie receptor kinase cDNA.	3845	128	455	7.07	0
41.	Q46081	Sequence derived from human c	2726	127	449	7.01	0
42.	Q53993	Myotonic dystrophy (DM) gene	3407	127	449	7.01	0
		**** 6 standard deviations above mean ****					
43.	Q06845	Sequence encoding tumour grow	2530	126	422	6.94	0
44.	Q03259	Pseudorabies virus glycoprote	4897	126	424	6.94	0
45.	Q42383	TGF-beta1 receptor type II (c	2090	125	270	6.88	0

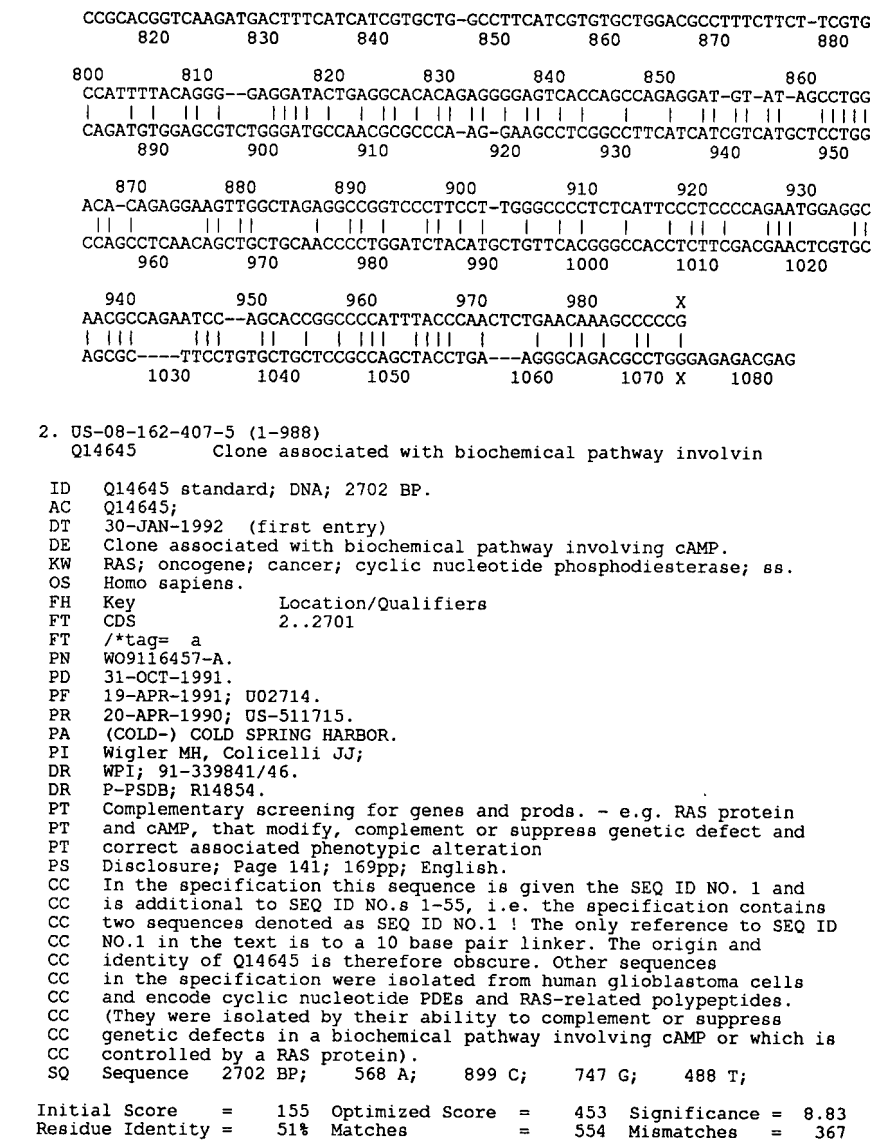
1. US-08-162-407-5 (1-988)
   Q41760 Oxytocin receptor coding sequence.

ID Q41760 standard; DNA; 1167 BP.
AC Q41760;
DT 02-SEP-1993 (first entry)
DE Oxytocin receptor coding sequence.
KW Human uterine myometrium tissue; labour; hypertonic uterine
KW dysfunction; ss.
OS Homo sapiens.
PN EP-542424-A.
PD 19-MAY-1993.
PF 09-OCT-1992; 309239.
PR 09-OCT-1991; JP-262167.
PR 27-MAR-1992; JP-071258.
PA (ROHT) ROHTO PHARM CO LTD.
PI Azuma C, Kimura T, Okayama H, Saji F, Tanizawa O;
DR WPI; 93-160903/20.
DR P-PSDB; R37264.
PT Isolated oxytocin receptor and DNA encoding it - obtd. from human
PT uterine myometrium tissue, used in study and treatments during
PT labour
PS Claim 8; Page 9-11; 15pp; English.
CC RNA was isolated from human myometrium tissue and used to prep. a
CC cDNA library in E. coli DH5 cells. Plasmid DNA was isolated,
CC linearised and used to synthesise mRNA transcripts. The mRNA was
CC injected into Xenopus oocytes and membrane currents after perfusion
CC with oxytocin were examined. Positive pools were selected and the
CC cDNA encoding the oxytocin receptor isolated. The oxytocin receptor
CC can be used to determine oxytocin receptor levels, partic. in
CC uterine endometrium cells. This can be used to e.g. infer labour
CC time, predict whether or not relief of labour pains of patients with
CC preterm labour is possible or for prevention of hypertonic uterine
CC dysfunction or determination of the time of inducing labour pains.
SQ Sequence 1167 BP; 178 A; 404 C; 363 G; 222 T;

Initial Score = 193 Optimized Score = 455 Significance = 11.29
Residue Identity = 51% Matches = 539 Mismatches = 394
Gaps = 120 Conservative Substitutions = 0







2. US-08-162-407-5 (1-988)
Q14645 Clone associated with biochemical pathway involvin

ID Q14645 standard; DNA; 2702 BP.
AC Q14645;
DT 30-JAN-1992 (first entry)
DE Clone associated with biochemical pathway involving cAMP.
KW RAS; oncogene; cancer; cyclic nucleotide phosphodiesterase; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 2..2701
FT /*tag= a
PN WO9116457-A.
PD 31-OCT-1991.
PF 19-APR-1991; U02714.
PR 20-APR-1990; US-511715.
PA (COLD-) COLD SPRING HARBOR.
PI Wigler MH, Colicelli JJ;
DR WPI; 91-339841/46.
DR P-PSDB; R14854.
PT Complementary screening for genes and prods. - e.g. RAS protein
PT and cAMP, that modify, complement or suppress genetic defect and
PT correct associated phenotypic alteration
PS Disclosure; Page 141; 169pp; English.
CC In the specification this sequence is given the SEQ ID NO. 1 and
CC is additional to SEQ ID NO.s 1-55, i.e. the specification contains
CC two sequences denoted as SEQ ID NO.1 ! The only reference to SEQ ID
CC NO.1 in the text is to a 10 base pair linker. The origin and
CC identity of Q14645 is therefore obscure. Other sequences
CC in the specification were isolated from human glioblastoma cells
CC and encode cyclic nucleotide PDEs and RAS-related polypeptides.
CC (They were isolated by their ability to complement or suppress
CC genetic defects in a biochemical pathway involving cAMP or which is
CC controlled by a RAS protein).
SQ Sequence 2702 BP; 568 A; 899 C; 747 G; 488 T;

Initial Score = 155 Optimized Score = 453 Significance = 8.83
Residue Identity = 51% Matches = 554 Mismatches = 367

```
Gaps            =    153  Conservative Substitutions            =     0

           X       10        20        30        40            50
           CGGCCGGAATTCCGGGGCCCCCGGCCGAAATGACAGTGCTGGC----GCCAGCCT---GGA-
           | || |   | | |  | || |   || | || | |  |  ||    |||| |||   |||
AGCCTGCACGCAGCTGACGTGCTGCAGTCCACCCACGTACTG-CTGGCCACGCCTTGGCCAACCTTAAGGAA
    1230 X   1240      1250      1260      1270      1280      1290

          60            70        80        90       100           110
-GC-----CCAACAACCT----AT-CT--CCTCCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGG--GACCC
 ||      ||   ||||    || ||  || ||  || | || ||||   ||   ||   ||||   ||||
TGCAGTGTTCACGGACCTGGAGATTCTCGCCGCCCTCTTC-GCGGCTGCCATC--CACGATGTGGATCACCC
   1300      1310      1320      1330      1340      1350      1360

      120       130       140       150       160       170
AGGACTGCTCCTTCCAACACAG--CCCCATCTCCTCCGACTTC---GCTGTCAAAATCCGTGAGCTGTCTGA
 ||   | ||  |||||| |||  || |||| | || | |||   |||| |    ||    |      |||
TGG---GGTC--TCCAAC-CAGTTCCTCATCAACACC-AATTCGGAGCTGGC--GCTCATGTACAACGATGA
         1370      1380      1390      1400      1410      1420

180       190       200       210       220       230       240       250
CTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGGCCTCTG
 |   ||||  |  || |||   |  |||||| || ||| |||| | | |.||||   ||||   | | |||
GTCGGTGCTCGAGAATCACCACCTGGCCGTGGGCTTCAAGCTGCTGCAGGAGGACAACTGC---GACATCTT
 1430      1440      1450      1460      1470      1480      1490

       260       270       280       290       300       310
GCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCT---CAAGACTGTC----GCTGGGTCCAAGATGCAAG
 | |    ||   ||| |||||  |  | || | ||   |||||  |||     |  ||| | |   || |
CCAGAACCTC--AGCA-AGCGCCAGCGGCAGAGCCTACGCAAGATGGTCATCGACATGGTGCTGG--CCACG
  1500      1510      1520      1530      1540      1550      1560

   320       330       340       350       360       370       380
G-CTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTC-AGCCCCCCCCCAGCT-
| | ||    ||| | ||| |     |    || |    ||| |  |||      | ||       ||||||
GACATGTCCAAGCACATGACCCTCCTGGCTGACCTGAAGACC-ATGGTGGAGACCAAGAAAGTGACCAGCTC
     1570      1580      1590      1600      1610      1620      1630

    390       400       410       420       430       440       450
-GTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAG
 |   || || |  | |||  |||  |||| || | | |||  | |||||| | |  |||| || |||  |
AGGGGTC-CTCCTGCTAGA-TAACTACTCCGACCGCATCCAG--GTCCTCCG-GAACATGGT-GCACTG-TG
     1640      1650      1660      1670      1680      1690

  460       470       480       490       500       510       520
CCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGAC--TCCTCAACCCTGC
||   || | |  ||   |||  | ||  |  |||||||| |  | |||   |||   || || |   ||
CC---GACCTC-AGC---AACCCCACCAAGCCGCTGGAGCTGTACCGCCAG-TGGACAGACCGCATCATGGC
  1700      1710      1720      1730      1740      1750      1760

  530       540       550       560       570       580       590
CACCCCCATGGAGTCCCCGGCCCCTGGAGGCCACAGCCCCGACAGCCCCGCAGCCCCCTCTGCTCC--TCCT
|     | |  || |   |   ||  || ||   |||   |   |     ||||||| | ||   |   |
CGAGTTCTTCCAG-CAGGGTGACCGAGA-GCGCGAGCGTGGCATGGAAATCAGCCCCATGTGTGACAAGCAC
     1770      1780      1790      1800      1810      1820      1830

  600       610               620       630       640       650       660
ACTGCTGCTGCCCGTGG----GCCTC---CTGCTGCTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGC
|||||  ||  ||||||    | |||     | | |     || | | ||||   || ||   |||  |
ACTGC--CT--CCGTGGAGAAGTCTCAGGTGGGTTTTATTGACTACATTGTGCACCCATTGTGGGAGACCTG
```



```
          1840      1850      1860      1870      1880      1890

            670           680       690       700       710       720
   GGCGGA----GGACACCC------CGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTG
   ||||||    |  |||||       |||| |  |  ||| | ||||  |  | |||  ||| ||||   | |
   GGCGGACCTTGTCCACCCATACAGCGCCATCCG--GCA-GAGCCCATCTCCGCCA--CCCGAGGAGGAG-TC
1900      1910      1920      1930          1940      1950          1960

        730           740       750       760       770       780
   CTTGTGG--AG--CACTGA--CCTG---GCCAAGGCCTCATCCTGCGGAG-CCTTAAACAACGCAG-TGAGA
      | ||  ||  |||  |  ||||   | ||||  | ||   || |  | |  |  |  | | ||  |||
   AAGGGGGCCAGGCCACCCACCCCTGCCTGACAAGTTC-CAGTTTGAGCTGACGCTGGAGGAGGAAGAGGAGG
    1970      1980      1990      2000      2010      2020      2030

      790           800       810       820       830           840
   CAGACATCTATCA---TCCCATTTTACAGGGGA-GGATACTGAGGCA--CACAGAG--GGGA--GTCACCAG
    ||| |  |||||    ||||   ||| | | |  |     ||||||   || |||  ||||  |||| ||
   AAGAAA--TATCAATGGCCCA-GATACCGTGCACAGCCCAAGAGGCATTGACTGAGCAGGGATTGTCAGGAG
   2040        2050      2060      2070      2080      2090      2100

     850       860       870       880       890       900       910
   CCAGAGGATGTATAGC-CTGGACACAGAGGAAGTTGGCT-AGAGGCCGGTCCCTTCCTTGGGCCCCTCTCAT
    | |||||     ||| |||||  || |  |  | | ||  |||||    ||   ||  | | |  |   |
   TC-GAGGA-----AGCTCTGGATGCA-ACCA--TAGCCTGGGAGGCATCCCCGGCCCAGGAGTCGTTGGAAG
     2110          2120          2130      2140      2150      2160

     920       930       940       950       960       970       980
   TCCCTCCCCAGAATGGAGGCAACGCCAGAATCCAGCACCGGCCCCA-TTTACCCAACTCTGAA--CAAAGCC
   |     | ||    ||| ||| | || | |  | |  | ||   || | ||    || | | |  || || |
   TTATGGCACA----GGAAGCATC-CCTGGAGGCCGAGCTGGAGGCAGTGTATTTGACACAGCAGGCACAGTC
2170          2180      2190      2200      2210      2220      2230

    X
   CCCG
   | |
   CACAGGCAGTGCAC
    2240
```

3. US-08-162-407-5 (1-988)
   Q05536 Sequence enoding purified soluble intercellular ad

```
ID   Q05536 standard; DNA; 1443 BP.
AC   Q05536;
DT   07-DEC-1990  (first entry)
DE   Sequence enoding purified soluble intercellular adhesion molecle-1
DE   (sICAM-1).
KW   Soluble intercellular adhesion molecule-1; HeLa; LFA-1; rhinovirus;
KW   Coxsackie A virus; ds;
OS   Homo sapiens.
PN   EP-379904-A.
PD   01-AUG-1990.
PF   12-JAN-1990; 100557.
PR   24-JAN-1989; US-301192.
PR   13-DEC-1989; US-445951.
PA   (MOLE-) MOLECULAR THERAPEUT.
PI   Greve J, Mcclelland A;
DR   WPI; 90-232770/31.
DR   P-PSDB; R06240.
PT   Human soluble inter-cellular adhesion molecule-1 - which inhibits
```

```
PT   lymphocyte function associated with LFA-1 and ICAM-1 interaction
PS   Claim 2; Fig 1; 16pp; English.
CC   sICAM can inhibit lymphocyte function associated with antigen
CC   LFA-1 and ICAM-1 interaction, useful in treating inflamation,
CC   graft rejection, and for LFA-1 expressing tumour cells. Can also
CC   inhibit infection by rhinovirus and Coxsackie A virus, increase
CC   cell adhesion and reverse immune dysfunction due to excess sICAM-1.
SQ   Sequence   1443 BP;    306 A;    453 C;    433 G;    251 T;

Initial Score     =    150  Optimized Score  =    453  Significance =  8.50
Residue Identity  =    50%  Matches         =    542  Mismatches   =   395
Gaps              =    126  Conservative Substitutions              =     0
```

```
            X         10          20          30          40          50
            CGGCCGGAA--TTCCG---GGGCCCCCGGCCGAAATGACAGTG-CTGGCGCCAG-CCTGGAG
AGTCATCCTGCCCCGGGGAGGCTCCGTGCTGGTGACATGCAGCACCTCCTGTGACCAGCCCAAGTTGTTGGG
   110    X  120         130         140         150         160         170

       60          70          80          90          100         110         120
CCCA-ACAACCTATCTCCT--CCTGCTGCTGCTGCTGAGCTCGGGAC--TCAGTGGGACCCAGGACT--GCT
CATAGAGACCCCGTTGCCTAAAAAGGAGTTGCTCCTG-CCTGGGAACAACCGGAAGGTGTATGAACTGAGCA
 180         190         200         210         220         230         240

        130         140         150         160         170         180         190
CCTTCCAACACAGCCCCATCTCCTCCGACTTCGCT-GTCAAAAT-CCGTGAGCTGTCTGACTACCTGCTTCA
ATGTGCAAGA-AG-----ATAGCCAACCAATGTGCTATTCAAACTGCCCTGA-TGGGCAGTC-AACAGCTAAA
250             260         270         280         290         300         310

        200         210         220         230             240         250
AGATTACCCA--GTCACCGTGGCCTCCA-ACCTGCAGG-ACGAGGAGCTCTGC------GGGGGCCTCTGGC
ACCTTCCTCACCGTGTAC-TGGACTCCAGAACGGGTGGAACTGGCACCCCTCCCCTCTTGGCAGCCAGTGG-
      320         330         340         350         360         370         380

      260         270         280         290         300         310         320
GGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGG
GCAAGAACCTTACCCTACGCTGCCAGGTGGAGGGTGGGGC--ACCCCGGG-CCAACCTCACCGTGGTGCTGC
      390         400         410         420         430         440         450

    330         340         350         360             370         380
AGCGCGTGAACACGGAGAT-ACACTTTGTCACCAAATGTG--------CCTTTCAGCCCCCCCCCAGCTGTCT
TCCGTGGGGAGAAGGAGCTGAAAC--GGGAGCCAGCTGTGGGGGAGCCCGCTGAGGTCACGACCA-CGGTGC
       460         470         480         490         500         510         520

390         400         410         420         430         440         450
TCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGAC-CTCCGAGCAG--CTGGTGGCGCTGAAGCC
T-GGTGAGGAGAGATCACCA--TGGAGCCAATTTCTCGTGCCGCACTGAACCTGGACCTGCGGC-CCCAAGGG
          530         540         550         560         570         580

460         470         480         490         500         510         520
CTGGATCACTCGCCAGAACTTCTCCCGGTGCCT-GGAGCTGCAG-TGTCAGCCCGACTCCTCAACCCTGCCA
CTGGAGCTGT-TTGAGAACACCT-CGGCCCCCTACCAGCTCCAGACCTTTGTCCTGC-CAGCGACTCCCCCA
590         600         610         620         630         640         650
```



```
530         540         550         560         570         580         590
CCCCCATGGAGTCCCCGGCCCCTGGAGG-CCACAGCCCCGACAGCCCCG-CAGCCCCCT---CTG-CTCCTC
CAACTTGTCAGCCCCCGGGTCCTAGAGGTGGACACGCAGGGGACCGTGGTCTGTTCCCTGGACGGGCTGTTC
660         670         680         690         700         710         720

    600         610         620         630         640         650         660
CTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCGG
CCAGTCTCGGAGGCCCAGGTCCACCTGGCACTGG--GGGACC-AGAGGTTGAAC-CCCACA-GTCAC--CTA
730         740         750         760         770         780         790

  670         680         690         700         710         720         730
AGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTC-CCCAGTCCCC--AGGACCTGCTGCTTGTGGAGCAC
TGGCAACGACTCCTTCTCGGCCAAG-G-CCTCAGTCAGTGTGACCGCAGAGGACGAG-GGCACCCAGCG-GC
  800         810         820         830         840         850         860

  740         750         760         770         780         790         800
TGACCTGGCCAAGGCCTCATCCTGCGGAGCCTTAAACA--ACGCAGTGAGACAGACATCTATCATC-CCATT
TGACGTGTGCA----GTAATACTGGGGAACCAGAGCCAGGAGACACTG---CAGACAGTGACCATCTACAGC
      870         880         890         900         910         920

    810             820         830             840             850
TTACAGGGG----AGG--ATACTGAGGCACACAGAGG-----GGAGTCACC-AG---CCAGAGGA-TGTATA
TTTCCGGCGCCCAACGTGATTCTGACGAAGCCAGAGGTCTGAGAAGGGACCGAGGTGACAGTGAAGTGTGAG
930         940         950         960         970         980         990

860         870         880         890             900         910         920
  GCCTGGACAC-AGAG-GAAGTTG--GCTAGAGGCCGGT----CCCTTCCTTGGGCCCCTCTCATTCCCTCCC
  GCC--CACCCTAGAGCCAAGGTGACGCTGAATGGGGTTCCAGCCCAGCCACTGGGCCC-GAGGGCCCAGCTC
1000        1010        1020        1030        1040        1050        1060

      930         940         950         960         970         980         X
  CAGAATGGAGGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCAACTCTGAACAAAGCCCCCG
  CTG-CTGAAGGCCACCCCAGAGGACA--ACGGGCGCAGCTTCTCCTGCTCTG--CAA--CCCTGGAGGTGGC
1070        1080        1090        1100        1110        1120        1130

  CGG
```

```
4. US-08-162-407-5 (1-988)
   Q21171       Clone encoding Intercellular Adhesion Molecule-1.

ID   Q21171 standard; DNA; 1846 BP.
AC   Q21171;
DT   21-MAY-1992  (first entry)
DE   Clone encoding Intercellular Adhesion Molecule-1.
KW   cloning technique; cell surface antigen; immunodiagnosis; ICAM-1;
KW   integrin; lymphocyte function associated antigen; LFA-1; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   sig_peptide     13..87
FT   /*tag=  a
```

```
FT   mat_peptide     88..1608
FT   /*tag= b
FT   /product= ICAM-1
PN   WO9201049-A.
PD   23-JAN-1992.
PF   15-JUL-1990; U04986.
PR   13-JUL-1990; US-553759.
PA   (GEHO-) GEN HOSPITAL CORP.
PI   Seed B, Aruffo A, Amiot M;
DR   WPI; 92-056864/07.
DR   P-PSDB; R20809.
PT   New CD53 cell surface antigen and DNA encoding it - for
PT   immuno-therapy and diagnosis of haematopoietic neoplasms, etc.
PS   Example 7; Fig 11; 160pp; English.
CC   A cDNA library was constructed using RNA prepared from HL60 cells
CC   induced with PMA. The library was transfected into COS cells  and
CC   cells expressing surface antigens were recovered by panning with
CC   the anti-ICAM MAbs 8F5 and 84H10. Episomal DNA was recovered from
CC   the panned cells and the expression-panning cycle repeated twice
CC   more to obtain a cDNA clone designated pICAM-1. The availability of
CC   a functional ICAM-1 cDNA will allow a better assessment of the role
CC   of ICAM-1/LFA-1 mediated adhesion in antigen-specific leukocyte
CC   function, including T-cell mediated killing, T-helper responses and
CC   antibody-dependent cell mediated killing.
SQ   Sequence   1846 BP;    416 A;    580 C;    531 G;    319 T;

Initial Score    =    149  Optimized Score  =    454  Significance =  8.44
Residue Identity =    50%  Matches          =    535  Mismatches   =   409
Gaps             =    112  Conservative Substitutions             =     0

            X         10         20         30         40         50
            CGGCCGGAA--TTCCG---GGGCCCCCGGCCGAAATGACAGTG-CTGGCGCCAG-CCTGGAG
            |  | || |   ||||    ||   |  || | |    | ||| |  || | ||   | | |
   AGTCATCCTGCCCCGGGGAGGCTCCGTGCTGGTGACATGCAGCACCTCCTGTGACCAGCCCAAGTTGTTGGG
    120        130        140        150        160        170        180

       60         70         80         90        100        110        120
   CCCA-ACAACCTATCTCCT--CCTGCTGCTGCTGCTGAGCTCGGGAC--TCAGTGGGACCCAGGACT--GCT
   |  | | | ||  |  |||     |  | |||| |||  || || ||   | |  ||     | |||  ||
   CATAGAGACCCCGTTGCCTAAAAAGGAGTTGCTCCTG-CCTGGGAACAACCGGAAGGTGTATGAACTGAGCA
 190        200        210        220        230        240        250

          130        140        150        160        170        180        190
   CCTTCCAACACAGCCCCATCTCCTCCGACTTCGCT-GTCAAAAT-CCGTGAGCTGTCTGACTACCTGCTTCA
      | ||| | ||    ||  ||  | | |  |||  ||||| | || |||   | | | | | | |||  |
   ATGTGCAAGA-AG----ATAGCCAACCAATGTGCTATTCAAACTGCCCTGA-TGGGCAGTC-AACAGCTAAA
 260        270            280        290        300        310        320

          200        210        220        230            240        250
   AGATTACCCA--GTCACCGTGGCCTCCA-ACCTGCAGG-ACGAGGAGCTCTGC------GGGGGCCTCTGGC
   |  || | ||  ||   | ||| ||||| | | |  || ||  | | | || |      ||  |||  |||
   ACCTTCCTCACCGTGTAC-TGGACTCCAGAACGGGTGGAACTGGCACCCCTCCCCTCTTGGCAGCCAGTGG-
      330        340        350        360        370        380        390

          260        270        280        290        300        310        320
   GGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGG
   |   | |||  | |  |||||   || |  |      | |   | | ||| |||  |    |   |||||
   GCAAGAACCTTACCCTACGCTGCCAGGTGGAGGGTGGGGC--ACCCCGGG-CCAACCTCACCGTGGTGCTGC
      400        410        420        430        440        450        460
```

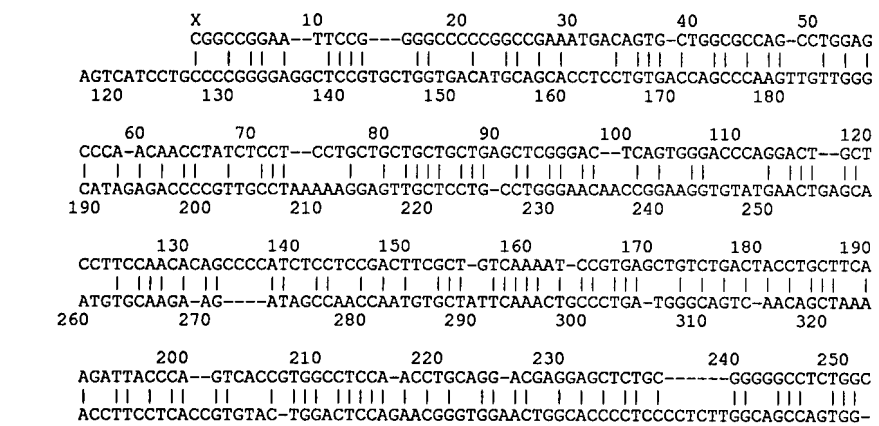


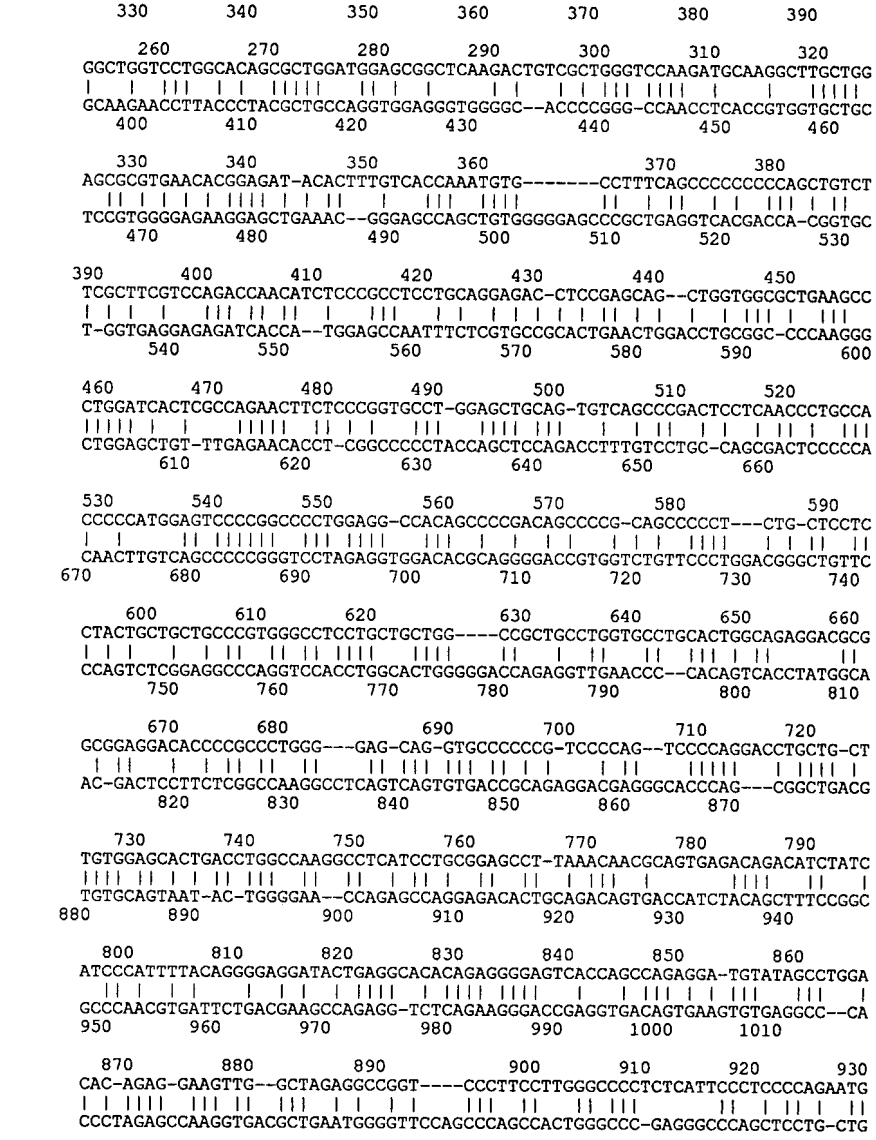
```
      330        340        350        360                370        380
   AGCGCGTGAACACGGAGAT-ACACTTTGTCACCAAATGTG-------CCTTTCAGCCCCCCCCCAGCTGTCT
     || | | | | |||| | | ||    |  |||  ||||        ||  | ||  | |  ||| | ||
   TCCGTGGGGAGAAGGAGCTGAAAC--GGGAGCCAGCTGTGGGGGAGCCCGCTGAGGTCACGACCA-CGGTGC
      470        480        490        500        510        520        530

  390        400        410        420        430        440        450
   TCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGAC-CTCCGAGCAG--CTGGTGGCGCTGAAGCC
   | | |  |  ||| || ||  |  |||   | |  | | | | | || | |  |  | ||| |  |||
   T-GGTGAGGAGAGATCACCA--TGGAGCCAATTTCTCGTGCCGCACTGAACTGGACCTGCGGC-CCCAAGGG
        540        550        560        570        580        590        600

   460        470        480        490        500        510        520
   CTGGATCACTCGCCAGAACTTCTCCCGGTGCCT-GGAGCTGCAG-TGTCAGCCCGACTCCTCAACCCTGCCA
   ||||| |  |    |||||  || | |   |||  |||| |||   |  | ||  | |  | || |  |||
   CTGGAGCTGT-TTGAGAACACCT-CGGCCCCCTACCAGCTCCAGACCTTTGTCCTGC-CAGCGACTCCCCCA
         610        620        630        640        650        660

   530        540        550        560        570        580           590
   CCCCCATGGAGTCCCCGGCCCCTGGAGG-CCACAGCCCCGACAGCCCCG-CAGCCCCCT---CTG-CTCCTC
   |  |     || ||||||  ||| ||||   |||  |  | | |   | | |  ||||   | | ||  ||
   CAACTTGTCAGCCCCCGGGTCCTAGAGGTGGACACGCAGGGGACCGTGGTCTGTTCCCTGGACGGGCTGTTC
670        680        690        700        710        720        730        740

      600        610        620            630        640        650        660
   CTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGG----CCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCG
   | | | -  |  | ||  || || ||||   ||||    ||   |  ||   ||  ||| | ||       ||
   CCAGTCTCGGAGGCCCAGGTCCACCTGGCACTGGGGGACCAGAGGTTGAACCC--CACAGTCACCTATGGCA
         750        760        770        780        790          800        810

         670        680            690        700        710        720
   GCGGAGGACACCCCGCCCTGGG---GAG-CAG-GTGCCCCCCG-TCCCCAG--TCCCCAGGACCTGCTG-CT
    | ||    |  | || ||  ||     || ||| ||| || | |    | ||    ||||   | |||| |
   AC-GACTCCTTCTCGGCCAAGGCCTCAGTCAGTGTGACCGCAGAGGACGAGGGCACCCAG---CGGCTGACG
         820        830        840        850        860        870

    730        740        750        760        770        780        790
   TGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCT-TAAACAACGCAGTGAGACAGACATCTATC
   |||| || | | || |||  ||  ||  | || |  ||  ||  | |||  |        ||||   ||  |
   TGTGCAGTAAT-AC-TGGGGAA--CCAGAGCCAGGAGACACTGCAGACAGTGACCATCTACAGCTTTCCGGC
880        890            900        910        920        930        940

   800        810        820        830        840        850        860
   ATCCCATTTTACAGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGGA-TGTATAGCCTGGA
     || |  | |    |  | |  | |||| | |||| ||||    |   | ||| | | |||   |||   |
   GCCCAACGTGATTCTGACGAAGCCAGAGG-TCTCAGAAGGGACCGAGGTGACAGTGAAGTGTGAGGCC--CA
   950        960        970        980        990       1000       1010

    870        880            890            900        910        920        930
   CAC-AGAG-GAAGTTG--GCTAGAGGCCGGT----CCCTTCCTTGGGCCCCTCTCATTCCCTCCCCAGAATG
   | | ||||  ||| ||  |||  | |  | |    |||  ||   || |||        ||  | || |  ||
   CCCTAGAGCCAAGGTGACGCTGAATGGGGTTCCAGCCCAGCCACTGGGCCC-GAGGGCCCAGCTCCTG-CTG
   1020       1030       1040       1050       1060       1070       1080

         940        950        960        970        980        X
   GAGGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCAACTCTGAACAAAGCCCCCG
    |||| || |||||   ||  || ||| |   ||  ||  |||||  |||  |||  |
   AAGGCCACCCCAGAGGACA--ACGGGCGCAGCTTCTCCTGCTCTG--CAA--CCCTGGAGGTGGCCGG
   1090       1100       1110       1120       1130       1140
```

5. US-08-162-407-5 (1-988)
N90611 Intercellular adhesion molecule-1 (ICAM-1) cDNA.

ID N90611 standard; cDNA; 1846 BP.
AC N90611;
DT 20-DEC-1989 (first entry).
DE Intercellular adhesion molecule-1 (ICAM-1) cDNA.
KW Tyrosine suppressor gene; ASV LTR; human cytomegalovirus AD169 enhancer;
KW HIV box; immunoselection; immune deficiency diseases; vasculitis;
KW systemic lupus erythematosus; rheumatoid arthritis; neoplasms;
FH Key Location/Qualifiers
FT CDS 13..1608
FT /*tag= a
PN EP-330191-A.
PD 30-AUG-1989.
PF 23-FEB-1989; 103127.
PR 25-FEB-1988; US-160416.
PA (GEHO) Gen. Hospital Corp.
PI Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo CP,
PI Simmons D, Stamenkovic I, Stengelin S;
DR WPI; 89-250302/35.
DR P-PSDB; P91357.
PT Rapid immuno:selection cloning - used to clone genes encoding
PT cell surface antigens associated with mammalian T lymphocytes.
PS Disclosure; fig. 11; 69pp; English.
CC ICAM-1 is a ligand for LFA-1 (a major receptor of T and B cells
CC and grnulocytes). ICAM contains no RGD motifs and is instead
CC homologous to the neural cell adhesion molecule (NCAM). COS cells
CC transfected with ICAM cDNA clone bind myeloid cells by a specific
CC interaction which can be blocked by monoclonal antibodies directed
CC against either ICAM-1 or LFA-1.
SQ Sequence 1846 BP; 416 A; 580 C; 531 G; 319 T;

Initial Score = 149 Optimized Score = 454 Significance = 8.44
Residue Identity = 50% Matches = 535 Mismatches = 409
Gaps = 112 Conservative Substitutions = 0





```
     1020       1030       1040       1050       1060       1070       1080

            940        950        960        970        980        X
     GAGGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCAACTCTGAACAAAGCCCCCG
      |||| || |||||   ||  || ||| |   ||  ||  |||||  |||  |||  |
     AAGGCCACCCCAGAGGACA--ACGGGCGCAGCTTCTCCTGCTCTG--CAA--CCCTGGAGGTGGCCGG
     1090       1100       1110       1120       1130       1140
```

6. US-08-162-407-5 (1-988)
   Q04049 Sequence of clone HL-60 encoding inter-cellular ad

```
 ID   Q04049 standard; cDNA; 3024 BP.
 AC   Q04049;
 DT   11-SEP-1990  (first entry)
 DE   Sequence of clone HL-60 encoding inter-cellular adhesion molecule
 DE   (ICAM-1).
 KW   Inter-cellular adhesion molecule; ICAM-1; Cancer;
 KW   metastasis; dexamethazone; azathioprine; cyclosporin A; ds.
 OS   Homo sapiens.
 FH   Key             Location/Qualifiers
 FT   CDS             58..1653
 FT   /*tag= a
 PN   WO9003400-A.
 PD   5-APR-1990.
 PF   28-SEP-1989; 004242.
 PR   28-SEP-1988; US-250446.
 PR   30-JUN-1989; US-373882.
 PA   (DANA-) Dana-Farber Cancer.
 PI   Springer T, Rothlein R, Marlin SD, Dustin ML;
 DR   WPI; 90-132247/17.
 DR   P-PSDB; R04165.
 PT   Inter-cellular adhesion molecules and binding ligands -
 PT   used for treating inflamation and tumour cells, etc.
 PS   Claim 1; Fig 8; 158pp; English.
 CC   Gene product may be used in treatment of inflamation caused by
 CC   deffence system response, and in supression of metastasis and
 CC   tumour growth. The product may also be used in conjunction with an
 CC   immunosuppressant selected from dexamethazone, azathioprine and
 CC   cyclosporin A.
 SQ   Sequence   3024 BP;   722 A;   874 C;   802 G;   626 T;   0 Others;

Initial Score    =    149  Optimized Score  =    454  Significance =  8.44
Residue Identity =    50%  Matches         =    534  Mismatches   =   412
Gaps             =    108  Conservative Substitutions             =     0

             X         10         20         30         40         50
             CGGCCGGAA--TTCCG---GGGCCCCCGGCCGAAATGACAGTG-CTGGCGCCAG-CCTGGAG
             |  | || |   ||||    ||   || | |    | ||| |  || | ||   | | |
     AGTCATCCTGCCCCGGGGAGGCTCCGTGCTGGTGACATGCAGCACCTCCTGTGACCAGCCCAAGTTGTTGGG
           170 X      180        190        200        210        220        230

         60         70         80         90        100        110        120
     CCCA-ACAACCTATCTCCT--CCTGCTGCTGCTGCTGAGCTCGGGACTCAGTGGGACCCAGGACTGCTCCTT
     |  | | | ||  |  |||     |  | |||| |||  || ||||  ||   |||    | | ||  | |
     CATAGAGACCCCGTTGCCTAAAAAGGAGTTGCTCCTG-CCT-GGGA-ACAACCGGAAGGTGTA-TGAACTTA
         240        250        260        270            280        290        300

          130        140        150        160        170        180        190
     CCAA--CACAGCCCCATCTCCTCCGACTTCGCT-GTCAAAAT-CCGTGAGCTGTCTGACTACCTGCTTCAAG
```



```
      |||     ||      ||  ||  | | |  |||  ||||| | || |||   | | | | | | |||  ||
     GCAATGTGCAAGAAGATAGCCAACCAATGTGCTATTCAAACTGCCCTGA-TGGGCAGTC-AACAGCTAAAAC
           310        320        330        340        350            360        370

          200        210        220        230                240        250
     ATTACCCA--GTCACCGTGGCCTCCA-ACCTGCAGG-ACGAGGAGCTCTGC------GGGGGCCTCTGGCGG
      || | ||  ||    | ||| ||||| | | |  || ||  | | | || |       ||  |||  ||| |
     CTTCCTCACCGTGTAC-TGGACTCCAGAACGGGTGGAACTGGCACCCCTCCCCTCTTGGCAGCCAGTGG-GC
           380        390        400        410        420        430        440

        260        270        280        290        300        310        320
     CTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGGAG
       |  |||  | |  |||||   || |  |      | |   | | ||| |||| |     |    ||||||
     AAGAACCTTACCCTACGCTGCCAGGTGGAGGGTGGGGC--ACCCCGGG-CCAACCTCACCGTGGTGCTGCTC
           450        460        470            480        490        500        510

      330        340        350        360                370        380        390
     CGCGTGAACACGGAGAT-ACACTTTGTCACCAAATGTG-------CCTTTCAGCCCCCCCCCAGCTGTCTTC
     || | | | |||| | | ||   |  |||  ||||         ||  | ||  | |  ||| | ||  ||  |
     CGTGGGGAGAAGGAGCTGAAAC--GGGAGCCAGCTGTGGGGGAGCCCGCTGAGGTCACGACCA-CGGTGCT-
           520        530        540        550        560        570

            400        410        420        430        440            450
     GCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGAC-CTCCGAGCAG--CTGGTGGCGCTGAAGCCCT
     | |  |   ||| || ||  |   |||   | |  | | | | | || | |  |  | ||| |  |||  ||
     GGTGAGGAGAGATCACCA--TGGAGCCAATTTCTCGTGCCGCACTGAACTGGACCTGCGGC-CCCAAGGGCT
    580        590        600        610        620        630        640

   460        470        480        490        500        510        520
     GGATCACTCGCCAGAACTTCTCCCGGTGCCT-GGAGCTGCAG-TGTCAGCCCGACTCCTCAACCCTGCCACC
     ||| |  |    |||| || | |  |||   |||| ||| |  | ||  | |  | || |  ||||
     GGAGCTGT-TTGAGAACACCT-CGGCCCCCTACCAGCTCCAGACCTTTGTCCTGC-CAGCGACTCCCCCACA
     650        660        670        680        690        700        710

   530        540        550        560        570        580            590
     CCCATGGAGTCCCCGGCCCCTGGAGG-CCACAGCCCCGACAGCCCCG-CAGCCCCCT---CTG-CTCCTCCT
      |     || ||||||  ||| |||   |||  |  |  | |  |  |||   | | ||  |||
     ACTTGTCAGCCCCCGGGTCCTAGAGGTGGACACGCAGGGGACCGTGGTCTGTTCCCTGGACGGGCTGTTCCC
      720        730        740        750        760        770        780

       600        610        620            630        640        650        660
     ACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGG----CCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGC
     | |   | | ||  || || ||||   ||||    ||  |  ||   ||  ||| | ||      ||  |
     AGTCTCGGAGGCCCAGGTCCACCTGGCACTGGGGGACCAGAGGTTGAACCC--CACAGTCACCTATGGCAAC
    790        800        810        820        830        840        850

        670        680            690        700        710        720
     GGAGGACACCCCGCCCTGGG---GAG-CAG-GTGCCCCCCG-TCCCCAG--TCCCCAGGACCTGCTG-CTTG
      ||  |  | || ||  ||     || ||| ||| || | |   | ||    |||||  | |||| | ||
     -GACTCCTTCTCGGCCAAGGCCTCAGTCAGTGTGACCGCAGAGGACGAGGGCACCCAG---CGGCTGACGTG
     860        870        880        890        900        910        920

      730        740        750        760        770        780        790
     TGGAGCACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCT-TAAACAACGCAGTGAGACAGACATCTATCAT
     || || | | || |||  ||  ||  | || |  ||  ||  | |||  |       ||||  ||  |
     TGCAGTAAT-AC-TGGGGAA--CCAGAGCCAGGAGACACTGCAGACAGTGACCATCTACAGCTTTCCGGCGC
      930        940        950        960        970        980        990

     800        810        820        830        840        850        860
     CCCATTTTACAGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGGA-TGTATAGCCTGGACA
```

```
 || |  | |    |  | |  | ||||  | |||| ||||   |   | ||| | | |||   |||   ||
CCAACGTGATTCTGACGAAGCCAGAGG-TCTCAGAAGGGACCGAGGTGACAGTGAAGTGTGAGGCC--CACC
  1000       1010       1020       1030       1040       1050       1060

870        880        890        900            910        920        930
C-AGAG-GAAGTTG--GCTAGAGGCCGGTCCCTTCCTTG----GGCCCCTCTCATTCCCTCCCCAGAATGGA
| ||||  ||| ||  ||| ||    ||| ||  ||  |     || || |     |||  | |    || |
CTAGAGCCAAGGTGACGCT-GAATGGGGTTCCAGCCCAGCCACTGCGCC-CGAGGGCCCAGCTCCTGCTGAA
   1070       1080       1090       1100       1110       1120       1130

     940        950        960        970        980        X
GGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCAACTCTGAACAAAGCCCCCG
||| || |||||   ||  || ||| |   ||  ||  |||||  |||  |||  |
GGCCACCCCAGAGGACA--ACGGGCGCAGCTTCTCCTGCTCTG--CAA--CCCTGGAGGTGGCCGG
   1140       1150       1160       1170          1180  X   1190
```

7. US-08-162-407-5 (1-988)
   N80995 HL-cDNA clone encoding human intercellular adhesio

```
 ID   N80995 standard; cDNA; 3113 BP.
 AC   N80995;
 DT   28-DEC-1990  (first entry)
 DE   HL-cDNA clone encoding human intercellular adhesion molecule-1
 DE   (ICAM-1)
 KW   Lymphocyte recognition; antiinflammatory agent;
 KW   lymphocyte function associated antigen 1 (LFA-1); ss.
 OS   Homo sapiens.
 FH   Key             Location/Qualifiers
 FT   sig_peptide     58..138
 FT   /*tag=  a
 FT   mat_peptide     139..1746
 FT   /*tag=  b
 FT   polyA_signal    2976..2981
 FT   /*tag=  c
 PN   AU8815518-A.
 PD   10-NOV-1988.
 PF   29-APR-1988; 106901.
 PR   04-MAY-1987; US-045963.
 PA   (DANA-) Dana Farber Cancer.
 PI   Springer TA, Rothlein R, Marlin SD, Dustin ML;
 DR   WPI; 89-033081/05.
 DR   P-PSDB; P80458.
 PT   New inter-cellular adhesion molecule-7, derived antibodies etc. -
 PT   are used for diagnosis and treatment of inflammation and
 PT   tumours, and for new DNA coding sequences
 PS   Claim 8; Fig 8; 74pp; English.
 CC   ICAM-1 and functional derivs. are new. Also new are recombinant DNA
 CC   able to express ICAM-1 or derivs. ICAM-1 is able to bind to a molecule
 CC   present on the surface of lymphocytes and contains at least one of 17
 CC   specified AA sequences (p80441 to p80457). ICAM-1 is a 76-97kD
 CC   glycoprotein which is a binding partner for LFA-1, and is involved in
 CC   lymphocyte recognition and adhering to cell surfaces. ICAM-1 contains
 CC   5 Ig-like domains.
 SQ   Sequence   3113 BP;   735 A;   905 C;   833 G;   640 T;
```

```
Initial Score    =   149  Optimized Score  =   451  Significance =  8.44
Residue Identity =   51%  Matches          =   542  Mismatches   =   390
Gaps             =   130  Conservative Substitutions             =     0
```



```
           X         10        20        30        40        50
           CGGCC--GGAATTCCGGGGCCCCCGGCCGAAATGA-CAGTGCTGGCGCCAGCCTGGAGCCCA
           | |||  ||||   |  ||      |  ||| ||| || || ||   | ||     |||| |
GGAGTTGCTCCTGCCTGGGAACAAC-CGGAAGGTGTATGAACTGAGCAATG-TG---CAAGAAGATAGCCAA
260        270       280       290       300           310       320

60          70        80        90        100           110
ACAA---CCTATCTCCTCCTG--CTGCTGCTGCTG-----AGCT----CGGGACTCA--GTG----GGAC-C
 |||    |||| ||   |||  ||| ||  || |     ||||    |    ||||  |||    |||| |
CCAATGTGCTAT-TCAAACTGCCCTGATG-GGCAGTCAACAGCTAAAACCTTCCTCACCGTGTACTGGACTC
   330       340       350       360       370       380       390

       120       130       140       150       160       170       180
CAGGACTGCTCCTTCCAACACAGC-CCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTCTGACTA
||| || | |    |  |||  | ||| |||   | | |   |  | ||| | || | | |  |  | ||
CAGAACGGGTGGAACTGGCACCCCTCCCCTCTTGGCAGCC--AGTGGGCAAGAACC-TTACC-CTACG-CTG
   400       410       420       430       440       450       460

      190       200       210       220       230           240
CCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCT-CTGC------GGGGGCC
|| | |  | | |    ||  | ||| |||   |||||| ||   || || ||| || |      ||  |||
CCAGGTGGAGGGTGGGGCA--CCCCG-GGC---CAACCT-CA--CCGTGGTGCTGCTCCCCTCTTGGCAGCC
      470       480             490       500       510       520

 250       260       270       280       290       300       310
TCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCT
  ||| |  ·|  ||| | |  |||||   || |  |     | |   | | ||| |||| |    |
AGTGG-GCAAGAACCTTACCCTACGCTGCCAGGTGGAGGGTGGGGC--ACCCCGGG-CCAACCTCACCGTGG
    530       540       550       560       570       580       590

320       330       340       350       360           370       380
TGCTGGAGCGCGTGAACACGGAGAT-ACACTTTGTCACCAAATGTG-------CCTTTCAGCCCCCCCCCAG
|||||   || | | | | |||| | | ||   |  |||  ||||        ||  | ||  | |  |||
TGCTGCTCCGTGGGGAGAAGGAGCTGAAAC--GGGAGCCAGCTGTGGGGGAGCCCGCTGAGGTCACGACCA-
     600       610       620       630       640       650       660

     390       400       410       420       430       440           450
CTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGAC-CTCCGAGCAG--CTGGTGGCGCT
| ||  | | |  |   ||| || ||  |   |||   | |  | | | | | || | |  |  | ||| |
CGGTGCT-GGTGAGGAGAGATCACCA--TGGAGCCAATTTCTCGTGCCGCACTGAACTGGACCTGCGGC-CC
       670       680       690       700       710       720

     460       470       480       490       500       510       520
GAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCT-GGAGCTGCAG-TGTCAGCCCGACTCCTCAACC
 |||  |||||  |  |    |||||  || | |   |||   |||| |||   |  | ||  | |  | ||
CAAGGGCTGGAGCTGT-TTGAGAACACCT-CGGCCCCCTACCAGCTCCAGACCTTTGTCCTGC-CAGCGACT
730       740       750       760       770       780       790

     530       540       550       560       570       580
CTGCCACCCCCATGGAGTCCCCGGCCCCTGGAGG-CCACAGCCCCGACAGCCCCG-CAGCCCCCT---CTG-
|  ||||  |     || ||||||  ||| ||||   |||  |  |  | |   | | |  ||||    | |
CCCCCACAACTTGTCAGCCCCCGGGTCCTAGAGGTGGACACGCAGGGGACCGTGGTCTGTTCCCTGGACGGG
800       810       820       830       840       850       860       870

 590       600       610       620           630       640       650
CTCCTCCTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGG----CCGCTGCCTGGTGCCTGCACTGGCAGAG
||  ||| | |   | | ||  || || ||||   ||||    ||   | ||  ||  ||| | ||
CTGTTCCCAGTCTCGGAGGCCCAGGTCCACCTGGCACTGGGGGACCAGAGGTTGAACCC--CACAGTCACCT
       880       890       900       910       920       930       940
```

```
      660       670       680            690       700       710
   GACGCGGCGGAGGACACCCCGCCCTGGG---GAG-CAG-GTGCCCCCCG-TCCCCAG--TCCCCAGGACCTG
      ||  | ||    |  | || ||  ||     || ||| ||| || | |    | ||    |||||  | |
   ATGGCAAC-GACTCCTTCTCGGCCAAGGCCTCAGTCAGTGTGACCGCAGAGGACGAGGGCACCCAG---CGG
           950       960       970       980       990      1000

  720       730       740       750       760       770       780
   CTG-CTTGTGGAGCACTGACCTGGCCAAGGCCTCATCCTGCGGAGCCT-TAAACAACGCAGTGAGACAGACA
   ||| | |||| || | | || |||  ||  ||  | || |  ||  ||  | |||  |      ||||
   CTGACGTGTGCAGTAAT-AC-TGGGGAA--CCAGAGCCAGGAGACACTGCAGACAGTGACCATCTACAGCTT
 1010      1020      1030      1040      1050      1060      1070

  790       800       810       820       830       840       850       860
   TCTATCATCCCATTTTACAGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAGGA-TGTATAG
   ||   | || |  | |   |  | |  | ||||  | |||| ||||   |   | ||| | | |||   |
   TCCGGCGCCCAACGTGATTCTGACGAAGCCAGAGG-TCTCAGAAGGGACCGAGGTGACAGTGAAGTGTGAGG
   1080      1090      1100      1110      1120      1130      1140

           870       880       890           900       910       920
   CCTGGACAC-AGAG-GAAGTTG--GCTAGAGGCCGGT----CCCTTCCTTGGGCCCCTCTCATTCCCTCCCC
   ||   || | ||||  ||| ||  |||  | |  | |    |||  ||   || |||       ||  | ||
   CC--CACCCTAGAGCCAAGGTGACGCTGAATGGGGTTCCAGCCCAGCCACTGGGCCC-GAGGGCCCAGCTCC
    1150      1160      1170      1180      1190      1200      1210

      930       940       950       960       970       980       X
   AGAATGGAGGCAACGCCAGAATCCAGCACCGGCCCCATTTACCCAACTCTGAACAAAGCCCCCG
    |  || |||| || |||||   ||  || ||| |   ||  ||  |||||  ||| |||  |
   TG-CTGAAGGCCACCCCAGAGGACA--ACGGGCGCAGCTTCTCCTGCTCTG--CAA--CCCTGGAGGTGGCC
    1220      1230      1240      1250      1260          1270  X   1280

   GG
```

8. US-08-162-407-5 (1-988)
   Q06583 Sequence encoding bovine metalloproteinase inhibit

```
ID   Q06583 standard; cDNA; 1045 BP.
AC   Q06583;
DT   21-FEB-1991  (first entry)
DE   Sequence encoding bovine metalloproteinase inhibitor.
KW   Tumour; chemotherapy; cancer; Paget's disease; osteoporosis;
KW   scleroderma; cholesteatoma; ds.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   CDS             289..948
FT   /*tag=  a
FT   mat_peptide     367..948
FT   /*tag=  b
PN   EP-398753-A.
PD   22-NOV-1990.
PF   18-MAY-1990; 305433.
PR   19-MAY-1989; US-355027.
PR   29-MAR-1990; US-501904.
PA   (AMGE-) AMGEN INC.
PA   (CHIL-) CHILDREN'S HOSPITAL OF LA.
PI   Langley KE, Boone TC, DeClerck YA;
DR   WPI; 90-350481/47.
DR   P-PSDB; R07954.
PT   New metallo-proteinase inhibitor polypeptide(s) - and DNA
```



```
PT   encoding them, for treatment of tumour cell dissemination and
PT   rheumatoid arthritis
PS   Claim 12; Fig 1; 63pp; English.
CC   Sequence may be used to transform a procaryotic or eukaryotic
CC   expression system to give a product with all the biological
CC   properties of naturally occuring metalloproteinase inhibitor.
CC   The product has therapuetic use in inhibiting tumour dissemination
CC   during chemotherapy and radiation therapy, impurged bone marrow cell
CC   harvesting etc. The inhibitor may also be useful in encapsulating
CC   tumours aiding clean excision, and in treatment of enphysema, Paget's
CC   disease, osteoporosis, scleroderma and bedsores.
CC   The gene product also has application in autoimmune disorders eg.
CC   rheumatoid arthritis and multiple sclerosis.
CC   See also Q06584.
SQ   Sequence   1045 BP;    219 A;    350 C;    311 G;    165 T;
```

```
Initial Score    =    143  Optimized Score  =    326  Significance =  8.05
Residue Identity =    50%  Matches          =    393  Mismatches   =   292
Gaps             =     95  Conservative Substitutions               =     0
```

```
     250       260       270       280       290       300       310
   CTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAG-GCT
              | | |||       ||  |||||||  |||  |||     | | ||||  |  ||  |  ||
             ATTCCGGC----TTCT--ATGGAGC-ACTCGGGAC-----CAG-GTCCGCGGCGCGCGCACT
             X           10          20          30          40

   320       330       340       350       360       370       380
   TGCTGGAGCGCGTGAACACGG--AGAT-ACACTT---TGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCT
    ||| |  |||       | | |  || |  | |||     | | |||    | |||    | || || || | |||
   CGCTCGCTCGCCGCCCCCCAGCCAGCTCTCGCTTCCGCGCCGCCAGCCGCGCC---CCGCGCCTCCTC-GCT
 50        60        70        80        90        100           110

      390       400       410       420       430       440       450
   G-TCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGA-GACCTCCGAGCAGCTGGTGGCGCTGAA
   |  |  |||  | |  || |||| |      |  |   |   || | ||   ||| ||   | | |   | |
   GCACCCCGCGACCTAGAG-CCAAGAAAGTTTGTGTGGCGAGTGAGGGCCGGAGAGGAGAGCGCGCCCGCGGA
   120       130       140       150       160       170       180

     460       470       480       490       500       510       520
   GCCCTGGATCACTCGCCAGAACTTCTCC--CGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTCAACCCTG
   |  | |  || |   |||| |  | || |||  |   ||||| ||   |  ||||| |  | |    | |
   GTGCCG--TC-CAGACCAGCGCGGCCCCGGCGGAGAGGGGAGC-GCCCCG--AGCCC-AGGCGGC--GGCGG
 190           200       210       220       230           240       250

     530         540       550         560       570       580
   CCACCCCCA--TGGAGTCCCCGGCCCCT--GGAGGCCACAGCCCCGACAGCCCCGCAGCC-CCCTCT-GC-T
   | | ||| |    | | |||| ||||||  |   ||||  | | |  | | |||||||||  || || || |
   CTAGCCCGAGTCCGCGACCCCCGCCCCTCCGCCCGCCATGGGCGCCGCCG-CCCGCAGCCTGCCGCTCGCGT
         260       270       280       290       300       310       320

 590       600       610       620       630       640           650
   CCT-CCTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCTGGTGC--CTGCACTG-GCA-GAGG
    || ||| |||||||||   | |   || || | | ||||  |||||  ||  || | | | |||   |
   TCTGCCTCCTGCTGCTG---GGGACGCTGCTCCCCCGGGCCGACGCCTGCAGCTGCTCCCCGGTGCACCCGC
        330           340       350       360       370       380       390

    660       670       680       690           700       710       720
   ACGCGGCG-GAGGACACCCCGCCCTGGGGAGCAGGTGCC--CCCCGTCCCCAGTCCCCAGGACCTGCTGCTT
   |   ||||    |  | | | | | || |||| |||    | ||   || |    ||||  || |  | |
   AACAGGCGTTTTGCAATGCAGACATAGTGATCAGG-GCCAAAGCAGT---CAATAAGAAGGAGGTG--GACT
```

```
       400       410       420       430       440       450

  730       740       750       760       770       780       790
GTGGAGCACTGACCTGGCCAAGG---CCTCATCCTGCGGAGCCTTAAACAACGCAGTGAGACAGACATCTAT
 ||  |||  ||| |  | | ||   || ||||  |||||  |   |  |   ||   ||| | | || | |
CTG--GCAACGACAT--CTACGGCAACCCCATCAAGCGGATTCAGTATGAGATCAAGCAGATAAAGATGT-T
460         470         480       490       500       510       520

   800       810       820       830       840       850       860
CATCCCATTTTA-CAGGGGAGGATACTGAGGCACACAGAGGGGAGTCACCAGCCAGAG-GATGTATAGCCTG
||    |  | | ||||  |       | |  ||||||       | | |   |  | | |  || | | |||
CAAGGGACCTGATCAGGACATAGAGTTTATCTACACAG-CCCCCGCCGCTGCCGTGTGTGGGGTCTCG-CTG
   530       540       550       560       570       580       590

    870       880           890       900           910       920
GACA-CAGAGG-AAGTTGG-----CTAGAGGCCGGTCCCTTCCTTGG----GCCCCTCT-CATTCCCTCCCC
||||   |||| |||  ||      ||    || ||       |   ||    | |  | | |||  |  || |
GACATTGGAGGAAAGAAGGAGTATCTCATTGCAGGGAAGGCCGAGGGGAATGGCAATATGCATATCACCCTC
   600       610       620       630       640       650       660

   930         940           950       960       970         980       X
AGAATGGAGGCAAC--GCC---AGAATCCAGCACCGGCCCCATTTACCCAACTCTGAA-CAAAGCCCCCG
 |  ||    || |  |||    ||  ||   |  | | ||    |    |  ||||| || ||   ||
TG--TGACTTCATCGTGCCCTGGGACACCCTGAGTGCCACCCAGAAGAAGAGCCTGAACCACAGGTACCAGA
   670       680       690       700       710       720       730   X

TGGGCTGT
 740
```

9. US-08-162-407-5 (1-988)
Q55132 Sequence encoding enzyme which desulphurises a fos

```
ID   Q55132 standard; DNA; 5535 BP.
AC   Q55132;
DT   02-AUG-1994  (first entry)
DE   Sequence encoding enzyme which desulphurises a fossil fuel.
KW   Enzyme; biocatalyst; fossil fuel; oxidation; cleavage;
KW   organosulphur compounds; coal; ds.
OS   Rhodococcus rhodochrous.
FH   Key             Location/Qualifiers
FT   CDS             2148..3245
FT   /*tag=  a
FT   /label= ORF 2
PN   WO9401563-A.
PD   20-JAN-1994.
PF   09-JUL-1993; U06497.
PR   10-JUL-1992; US-911845.
PA   (ENER-) ENERGY BIOSYSTEMS CORP.
PI   Denome SA,  Kovacevich BR,  Piddington CS,  Rambosek J;
PI   Young KD;
DR   WPI; 94-035068/04.
DR   P-PSDB; R47874.
PT   DNA encoding a bio catalyst which desulphurises fossil fuels -
PT   obtd. from Rhodococcus rhodochrous bacteria, used to produce
PT   microorganisms which degrade organic sulphur cpds.
PS   Claim 42; Page 75-79; 104pp; English.
CC   Microorganisms transformed with the DNA can be used to produce
CC   the enzyme/biocatalyst for the selective oxidative cleavage of
CC   carbon-sulphur bonds for desulphurisation of fossil fuels which
```



```
CC   contain organosulphur compounds.
SQ   Sequence   5535 BP;    1028 A;    1812 C;    1726 G;    969 T;

Initial Score    =    143  Optimized Score  =    449  Significance =  8.05
Residue Identity =    50%  Matches          =    535  Mismatches   =   399
Gaps             =    120  Conservative Substitutions               =     0

            X       10            20            30            40
            CGGCCGGAATTCCGGGGC-----CCC---CGGCCGAAA----TGACAGTGCTGGCGCCAG-C
              || ||     | ||||||      |||   | | |||||     | || |    ||  ||| |
  ATCAAAGCCGAGGTCGACGCTGCGGGGCGCGATCCCGATCAGACGAAAATCTTCACCGCCGTGATGCCGGTA
      1580  X   1590      1600      1610      1620      1630      1640

 50        60          70        80          90        100       110
  CTGGAGCCCAA--CAACCTATCTCCTCCTGCTGCTGCT-GCTGAGCTCGGGACTCAGTGGGACCCAGGACTG
  || | ||  ||  ||  |  |  |    | |   |  || |   | |||    | || |     | || | | ||
  CTCG-GCGAAAGCCAGGCGGTGGC--ACAGGAACGACTGGAATATCTCAACAGTCTGGTCCATCCGGAAGTG
     1650      1660        1670      1680      1690      1700      1710

 120       130       140       150         160       170       180
  CTCCTTCCAACACAGCCCCATCTCCTCCGACTTC--GCTGTCAAAATCCGTGAGCTGTCTGACTACCTGCTT
     ||  | || |   ||  ||  | ||| | ||   ||| |       |||   || || ||| ||     |
  GGACTGTCGACGCTATCCAGTC-ACACCGGCATCAACCTGGC----GGCGTACCCTCTC-GAC-ACTCCGAT
     1720      1730      1740      1750          1760      1770

 190       200       210         220       230       240       250
  CAAGATTACCCAGTCACCGTGGCCTCCA--ACCTGCAGGACGAGGAGC-TCTGCGGGGGCCTCTGGCGGCTG
  ||||   | || |     || |  || ||  | | | | | |  |  ||  ||||     |  ||  || || |
  CAAGGACATCCTG-----CGGGATCTGCAGGATCGGAATGTCCCGACGCAACTGCACATG-TTC--GCCGCCG
1780      1790          1800      1810      1820      1830      1840

 260       270       280         290       300       310       320
  GTCCTGGCACAGCGCTG-GATGGAGC-GGCTCAA--GACT-GTCGC--TGGGTCCAAGATGCAAGGC-TTGC
    |  ||||||||  | | |   || ||| ||  |  | | ||| |||  ||||  ||     | || |
  --CAACGCACAGCGAAGAGCTCACGCTGGCGGAAATGGGTCGGCGCTATGGAACCAACGTGGGGTTCGTTCC
      1850      1860      1870      1880      1890      1900      1910

        330       340       350       360           370       380
  T-GGAGCGCGTGAACACGGAGATACACTTTG-TCACCAAATGTGCC----TTTC-AG--CCCCCCCCCAGCT
  |  | | ||  | || ||| ||  ||  | | | || |  ||  ||      ||| ||  |  | || | | |
  TCAGTGGGCCGGTAC-CGGGGA-GCAGATGGCTGACGAGCTGATCCGCCACTTCGAGGGCGGCGCCGCGGAT
     1920      1930      1940      1950      1960      1970      1980

     390       400       410       420       430       440       450
  GTCTTC-GCTTCGTCCAGACCAACATCTCCCGCCTCCTGC-AGGAGACCTCCGAGCAGCTGGTGGCGCTGAA
  |  |||  | ||   | | ||  | |  || | ||||| | | |||  |  ||| ||| ||||  || |
  GGTTTCATCATCTCTCCGGCCTTCCT-GCCGGGCTCCTACGACGAGTTCGTCGACCAGGTGGTTCCGGTTCT
     1990      2000      2010      2020      2030      2040      2050

     460       470       480       490       500       510       520
  GCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCC
  |  | ||||| |  || |   | || || || || ||  ||   || || ||||  ||     ||||
  G--CAGGATCGC-GGCTACTTCCGCACCGAGTACCAGG-GCAACA--CTCTGCGCGACCACTTGGGTCTGCG
      2060      2070      2080      2090      2100      2110      2120

  530       540       550       560       570       580           590
  ACCCCCATGGAGTCCCCGGCC--CCTGGAGGCCACAGCCCCGACAGCCCCGCAGCCCCC---TCTG--CTC-
     |||   | | |  || |  |||  | | || |||| || |  |||||||  ||||   || |  |||
  CGTACCA-CAACTGCAAGGACAACCTTCATGACA-AGCCGCGTCGACCCCGCAAACCCCGGTTCAGAACTCG
        2130      2140      2150      2160      2170      2180      2190
```

```
          600       610       620       630       640       650
-CTCCTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCTGGTGC-CTGCAC-TGGCAGAGGACG
  |||  |  | ||  | |   | ||| | ||  ||| ||| | ||    || |||| |  ||||  |||
ATTCCGCCATCCGCGACACACTGACCTACAGCAACTGCCCGGTACCCAACGCTCTGCTCACGGCATCGGAAT
      2200      2210      2220      2230      2240      2250      2260

660       670       680       690       700          710       720
CGGCGGAGGACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCA--GTCC-CCAGGACCTGCTGCTTGT-
||  ||    | |  ||||   || | |  |  | |  ||||| ||  | ||  |||| |  | | | | |
CG--GGCTTCCTCGACGCCGCCGGCATC--GAACTCGACGTCCTCAGCGGCCAGCAGGGCACGGTTCATTTC
      2270      2280      2290      2300      2310      2320      2330

  730       740       750       760           770       780       790
--GGAGCACTGACCTGGCCAAGGCCTCATCCTGCG---GAG--CCTTAAACAACGCAGTGAGACAGACATCT
    | |||   ||| ||| |  || | |  || |   ||| ||    ||  | ||| |||   |    |
ACCTACCACCAGCCT-GCCTACACC-CGTTTTGGGGGTGAGATCCCGCCACTGCTCAGCGAG--GGGTTGCG
      2340      2350      2360      2370      2380      2390

     800       810       820       830       840       850       860
ATCATCCCATTTTACAGGGGAGGATACTGAGGCA-CACAGAGGGGAGTCACCAGCCAGAGGATGTATAGCCT
  ||  ||       ||   | |  |||| |||| |||     |     |     ||||| |  | | |  ||
GGCA--CCTGGGCGCA--CGCGTCTACT-CGGCATCAC-CCCGCTCTTGGGGCGCCAGGGCTTCTTTGTCCG
2400        2410        2420        2430        2440      2450      2460

    870       880       890       900       910       920       930
GGACACAGAGGAAGTTGGCTAGAGGCCGGTCCCTTCCTTG-GGCCCCTCTCATTCCCTCCCCAGAATGGAGG
 |||  | ||   | |  | || ||||  ||  |||||  |  | ||  ||    |  | | |  |   ||
CGAC-GACAGCCCGATCAC-AGCGGCCG---CCGACCTTGCCGGACGTCGAATCGGCGTCTCGGCCT--CGG
   2470      2480      2490        2500      2510      2520

   940       950       960       970       980       X
CAACGCCAGAATCCAGCACCGGCCCCATTTAC--CCAACTCTGAACAAAGCCCCCG
|||  |  | |||| ||  |||||   |   |  | | |||       | ||| ||
CAATTC--GCATCCTGC-GCGGCCAGCTGGGCGACTACCTCGAGTTGGATCCCTCGCGGCAAACGC
2530        2540      2550      2560      2570      2580 X    2590
```

10. US-08-162-407-5 (1-988)
    Q38735 Human OP-2.

```
ID   Q38735 standard; cDNA; 1723 BP.
AC   Q38735;
DT   15-JUL-1993  (first entry)
DE   Human OP-2.
KW   morphogenic; osteogenic protein; developmental cascade; hOP-2;
KW   inflammation; anti-inflammatory; Transforming Growth Factor;
KW   TGF-beta super-family; hippocampus; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             490..1698
FT   /*tag= a
FT   /note= "hOP2"
FT   mat_peptide     1279..1695
FT   /*tag= b
FT   /note= "contains conserved 7 cysteine skeleton"
PN   WO9304692-A.
PD   18-MAR-1993.
PF   28-AUG-1992; U07358.
PR   30-AUG-1991; US-752764.
```



```
PR   30-AUG-1991; US-752861.
PR   30-AUG-1991; US-753059.
PA   (CREA-) CREATIVE BIOMOLECULES INC.
PI   Cohen CM,  Kuberasampath T,  Oppermann H,  Ozkaynak E;
PI   Pang RHL, Rueger DC,  Smart JE;
DR   WPI; 93-100652/12.
DR   P-PSDB; R33410.
PT   Morphogen-induced modulation of inflammatory response - and
PT   resulting tissue damage, e.g. in autoimmune diseases, diabetes,
PT   asthma, ischemia reperfusion injury, etc.
PS   Claim 26; Page 119-121; 165pp; English.
CC   Human osteogenic protein (OP)-2 is a preferred morphogen for use in
CC   treating tissue damage in e.g. inflammatory disease, autoimmune
CC   disease, arthritis, psoriasis, dermatitis, diabetes and emphysema.
CC   Proteins having at least 70% homology with OP-2 amino acid
CC   sequences can also be used. See R33400 for mature hOP-2.
SQ   Sequence   1723 BP;    266 A;    625 C;    574 G;    258 T;

Initial Score    =   142  Optimized Score  =   393  Significance =  7.98
Residue Identity =   51%  Matches         =   474  Mismatches   =   333
Gaps             =   119  Conservative Substitutions             =     0

            X       10          20          30          40          50
            CGGCCGGAATTCCG-GGGCCCCCGGCCGAAAT---GACAGTGCTGGCG---CCAGC-CTGGA
             ||  ||| |||||       || | || | ||   | | | |  ||||    ||||  |||
   GGAGCCCCATTGGAAGGAGTTCCGCTTTGACCTGACCCAGATCCCGGCTGGGGAGGCGGTCACAGCTGCGGA
         860 X     870         880         890         900         910        920

            60          70          80          90          100         110
   G--CC-CA---ACAACCTATC--TCCTCCTGCTGCTGCTGCTGAGCTCGGGAC-TCAGTGGGACCCAGGACT
   |  ||  |   ||||  |  |   | |||  ||||| |  | |  | |   || ||||    |  |||||  |
   GTTCCGGATTTACAAGGTGCCCAGCATCCACCTGCT-CAACAGGACCCTCCACGTCAGCATGTTCCAGG--T
       930         940         950         960         970         980        990

   120         130         140         150         160         170         180
   GCTCCTTCCAACACAGCCCCATCTCCTCCGACTT-CGCTGTCAAAATCCGTGAGCTGTCTGACTACCTGCTT
   | |||     | ||   |  ||      || |||||   || |     || | | ||  | ||  | | |||
   GGTCC-AGGAGCAGTCCAACAGGGAGTCTGACTTGTTCTTTTTGGAT-CTTCAGACG-CT-CCGAGCTG-GA
        1000        1010        1020        1030        1040        1050

  190         200         210         220         230         240         250
   CAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAG-GAGCTCTGCGGGGGCCTCTGGCGGC---T
    | ||   |     |     |||    ||  | ||||  | || ||   |||| ||   ||   ||| |
   GACGAGGGCTGGCTGGTGCTGGATGTCA--CAGCAG--CCAGTGA---CTGCTGGTTGCTGAAGCGTCACAA
1060        1070        1080        1090        1100        1110        1120

    260         270         280         290         300         310         320
   GGTCCTGGCACAGCG-CTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGGCTTGCTGGAGC
   || ||||| ||  || ||  |||  | | || | ||| | | | | ||    ||| |  | || || ||
   GGACCTGGGACTCCGCCTCTATGTGGAGACTGAGGACGGGCACAGCGT---GGATCCTGGCCTGGCCGGCCT
     1130        1140        1150        1160        1170        1180        1190

   330         340         350         360         370         380         390
   GCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTC--GCTTCGT
   ||  |  ||  | |   |||  |  || ||      |||||||     |       || |||||  |    ||
   GCTGGGTCAACGGGCCCCACGCTCCCAACA-----GCCTTTCGTGGTC------ACTTTCTTCAGGGCCAGT
      1200        1210        1220          1230          1240        1250

    400         410         420         430         440         450         460
   CC-AGACCAACAT-CTCCCGCCTCCTGCAG-GAGACCTCCGAGCAGCTGG-TGGCGCTG--AAGC-CCTGGA
```

```
CCGAGTCC--CATCCGCACCCCTCGGGCAGTGAGGCCACTGAGGAGGAGGCAGCCGAAGAAAAGCAACGAGC
  1260        1270       1280       1290       1300       1310       1320

       470        480        490        500        510        520
T--CACTCGCCAGAACTTCTCCC--GGTGCCTGGAGCTGCAGT-GTCAGC-CCGAC--TCCTCAACCCTGCC
TGCCGCAGGCCA-ACCGACTCCCAGGGATCTTTGA--TGACGTCCACGGCTCCCACGGCCGGCAGGTCTGCC
  1330       1340       1350       1360       1370       1380       1390

 530        540        550        560        570        580        590
ACCCCCATGGAGTCCCCGGC-CCCTGGAGG-CCACAGC---CCCGACAGCCCCGCAGC-CCCCTCTGCTCCT
GTCGGCACGAGCTCTACGTCAGCTTCCAGGACCTCGGCTGGCTGGACTGGGTCATCGCTCCCCAAGCCTACT
   1400       1410       1420       1430       1440       1450       1460

     600        610        620        630        640        650        660
CCTACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCT-GCCTGGTGCCTGCACTGGCAGAGGACGCGGC
CGGCCTATTACTGTGAGGGGG---AGTGCTCCTTCCCACTGGACTCCTGCATG-AATGCCACCAACCACGCC
 1470       1480       1490       1500       1510       1520       1530

    670        680        690        700        710        720        730
GGAGGACACCCCGCCCTGG-GGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCAC
---ATCCTGCAGTCCCTGGTGCACCTGATG-AAGCCAAACGCAGTCCCCAAGGCGTGCTG--TG-----CAC
      1540       1550       1560       1570       1580       1590

   740        750        760        770        780        790        800
TGACCTGGCCAAGGCCTCATCCTGCGGAGC-CTTAAACAAC-GCAGTGAGACAGACATCTATCATCCCATTT
CCACCAAGCTGA-GCGCCA-CCT-CTGTGCTCTACTATGACAGCAGCAACAACGTCATC---CTGCGCA---
  1600       1610       1620       1630       1640       1650

   810        820          830        840        850        860        870
TACAGGGGAGGATACTG----AGGCACACAGAGGGGAGTCA-CCAGCCAGAGGATGTATAGCCTGGACACAG
AACACCGCAACATGGTGGTCAAGGCCTGCGGCTGCCACTGAGTCAGCCCGCCCAGCCCTA--CTG----CAG
1660       1670       1680       1690       1700       1710       1720      X

      880
AGGAAGTTGG
```

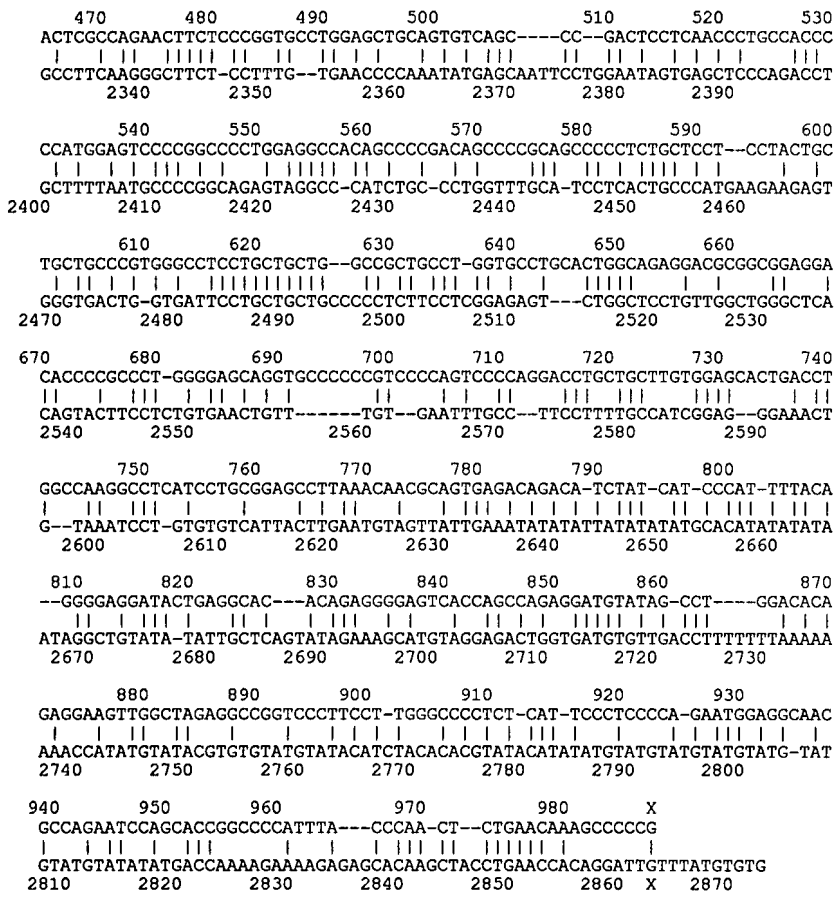
11. US-08-162-407-5 (1-988)
    Q01758 cDNA sequence of rat C kinase delta

```
ID   Q01758 standard; cDNA; 2909 BP.
AC   Q01758;
DT   03-AUG-1990  (first entry)
DE   cDNA sequence of rat C kinase delta
KW   Rat phosphorylating enzyme C delta; reagent; diagnostic chemical;
KW   rat C kinase delta.
OS   Rat.
FH   Key                 Location/Qualifiers
FT   CDS                 363..2384
FT   /tag= a
PN   J02000433-A.
PD   05-JAN-1990.
PF   08-OCT-1987; 252506.
```



```
PR   08-OCT-1987; JP-252506.
PA   (TAKE) Takeda Chemical Ind Kk.
DR   WPI; 90-047984/07.
DR   P-PSDB; R05228.
PT   Rat C kinase delta enzyme -
PT   expressed in animal cells and used as a reagent or diagnostic
PT   chemical
PS   Disclosure;Fig 1; 23pp; Japanese.
CC   It and its encoded rat phosphorylating enzyme C delta are new.  Also new
CC   are a transformant contg. a vector with it and prepn. of its polypeptide
CC   by culturing the transformant.  Its encoded polypeptide can be used as a
CC   reagent and a diagnostic chemical.
SQ   Sequence   2909 BP;   721 A;   773 C;   757 G;   658 T;
```

```
Initial Score    =    142  Optimized Score  =    428  Significance =  7.98
Residue Identity =    48%  Matches          =    514  Mismatches   =   417
Gaps             =    123  Conservative Substitutions              =     0

            X         10         20         30         40         50
            CGGCCGG--AATTCCGGGGC-CCCCGGCCGAAAT-GACAGTGCTGGCGCCAGCCTGGAGCCC
  GGGGAGAACCGGGCCAGCACATTCTGCGGCACTCCTGACTACATCGCCCCTG--AGATCCTG-CAGGGCCTG
  1860       1870       1880       1890       1900       1910       1920

  60         70         80         90         100        110        120
  AACAACCTATCTCCTCCTG--CTGCTG--CTGCTGAGCTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCC
  AAGTACTCATTTTC-CGTGGACTGGTGGTCTTTTGGGGTC-CTCCTCTATGAGA--------TGCTCATT-G
   1930       1940       1950       1960       1970               1980

   130        140        150        160        170        180        190
  AACACAGCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAG-CTGTCTGACTACCTGCTTCAAGATTAC
  GCCAGTCCCCCTTCCATGGTGATGATGAGGACGAGCTCTTTGAGTCCATCCGGGTGGACAC-ACCACACTAC
  1990       2000       2010       2020       2030       2040       2050

        200        210        220        230        240        250
  CC------AGTCACCGTGGCCTCCAA--CCTGCAGGACGAGGAGCTCT--GCGGGGGCCTCTGGCGGCTGGT
  CCGCGCTGGATCACCAAGGAGTCCAAGGACATCATG--GAGAAGCTCTTCGAGAGGGACCCTGCCAAGAGG-
 2060       2070       2080       2090       2100       2110       2120

260        270        280        290        300        310        320
  CCTGGCACAGCGCTGGATGGAGC-GGC-TCAAGACTGTCGC-TGGGTCCAAGATGCAAGGCTTGCTGGAG-C
  -CTGGGA-GTAACAGGAAACATCAGGCTTCACCCCTTTTTCAAGACTATCAACTGGAA--CCTGCTGGAGAA
  2130       2140       2150       2160       2170       2180       2190

  330        340        350        360        370        380        390
  GCGTGAACACGGAGATACACTTTGTCACCAAA-TG---TGCCTTTCAGCCCCCCCCCCAGCTGTCTTCGCTTC
  GCG-GAAGGTGGAGCCGCCCTTTAAGCCCAAAGTGAAATCCCCTTCAG-ACTACAGCAACTTTGACCCAGAG
     2200       2210       2220       2230       2240       2250       2260

    400        410        420        430        440        450        460
  GTCCAGACCAACATCTCCCGCC--TCCTGCAGGAGACCTCCGAGCAGCTGGT-GGCGCTGAAGCCCTGGATC
  TTCCTGAATGAGAAACCCCAACTTTCCTTCA-GTGA----CAAGAACCTCATCGACTCT-ATGGACCAGACA
    2270       2280       2290            2300       2310       2320       2330
```

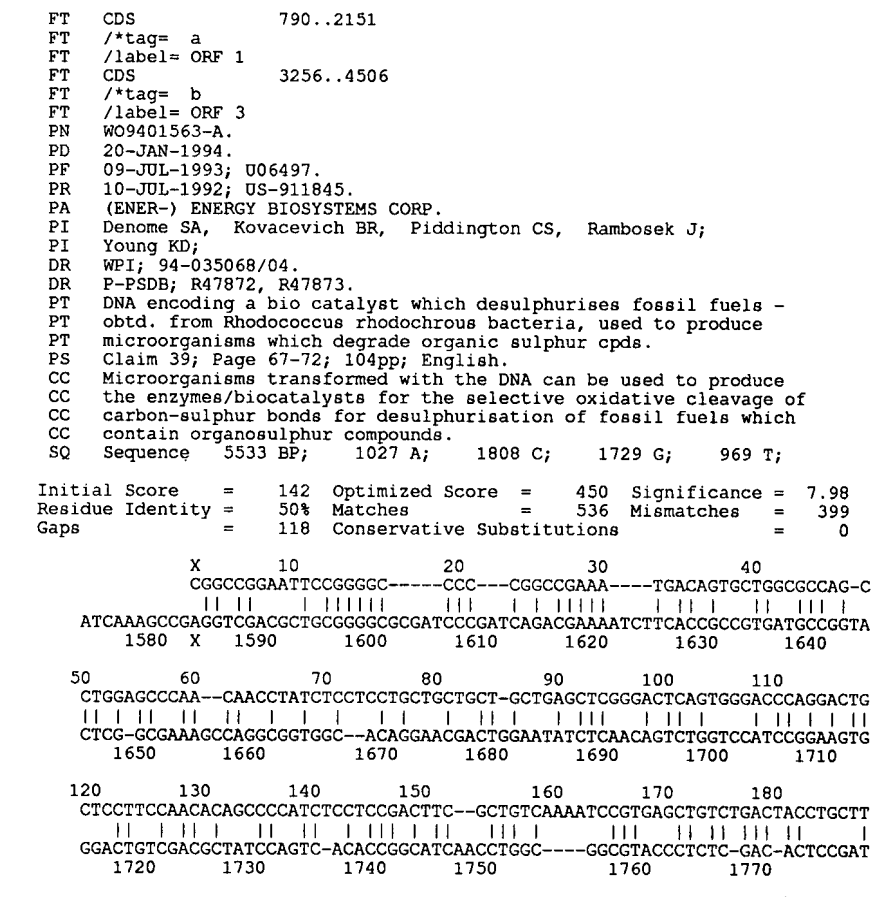


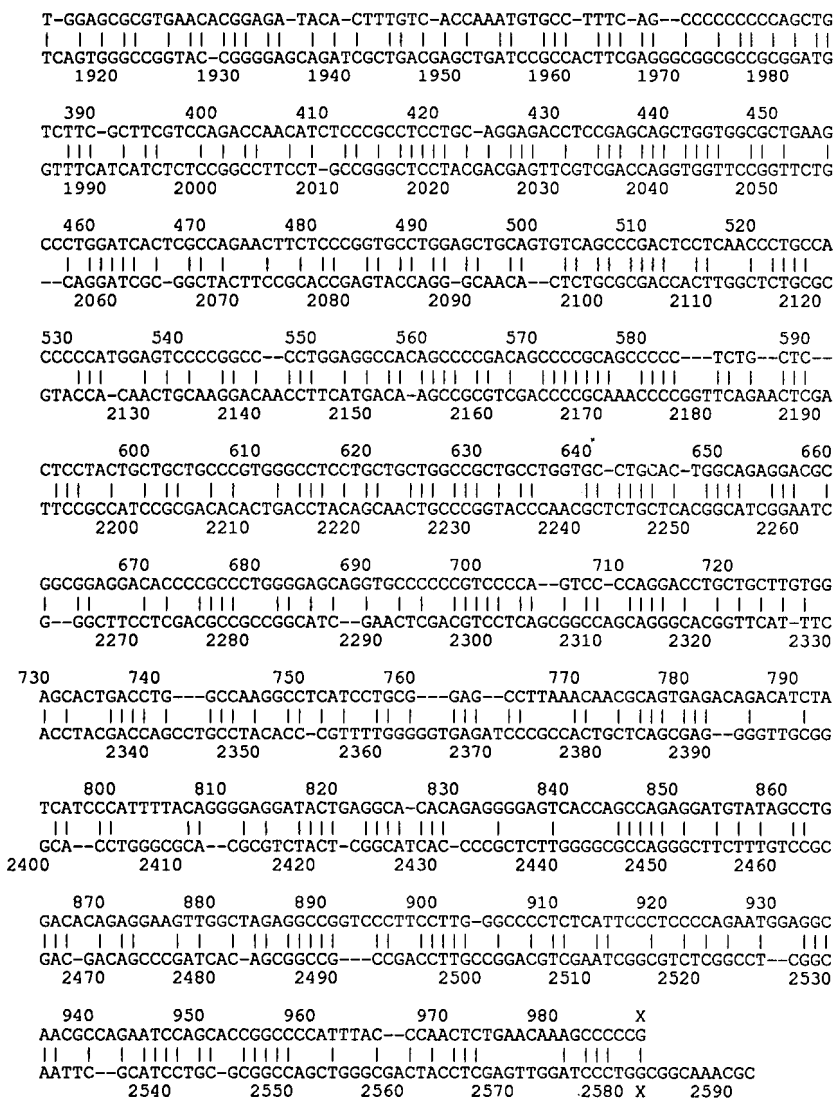
12. US-08-162-407-5 (1-988)
Q55131 Sequence encoding enzymes which desulphurise a fos

ID Q55131 standard; DNA; 5533 BP.
AC Q55131;
DT 02-AUG-1994 (first entry)
DE Sequence encoding enzymes which desulphurise a fossil fuel.
KW Enzyme; biocatalyst; fossil fuel; oxidation; cleavage;
KW organosulphur compounds; coal; ds.
OS Rhodococcus rhodochrous.
FH Key Location/Qualifiers

FT CDS 790..2151
FT /*tag= a
FT /label= ORF 1
FT CDS 3256..4506
FT /*tag= b
FT /label= ORF 3
PN WO9401563-A.
PD 20-JAN-1994.
PF 09-JUL-1993; U06497.
PR 10-JUL-1992; US-911845.
PA (ENER-) ENERGY BIOSYSTEMS CORP.
PI Denome SA, Kovacevich BR, Piddington CS, Rambosek J;
PI Young KD;
DR WPI; 94-035068/04.
DR P-PSDB; R47872, R47873.
PT DNA encoding a bio catalyst which desulphurises fossil fuels -
PT obtd. from Rhodococcus rhodochrous bacteria, used to produce
PT microorganisms which degrade organic sulphur cpds.
PS Claim 39; Page 67-72; 104pp; English.
CC Microorganisms transformed with the DNA can be used to produce
CC the enzymes/biocatalysts for the selective oxidative cleavage of
CC carbon-sulphur bonds for desulphurisation of fossil fuels which
CC contain organosulphur compounds.
SQ Sequence 5533 BP; 1027 A; 1808 C; 1729 G; 969 T;

Initial Score = 142 Optimized Score = 450 Significance = 7.98
Residue Identity = 50% Matches = 536 Mismatches = 399
Gaps = 118 Conservative Substitutions = 0

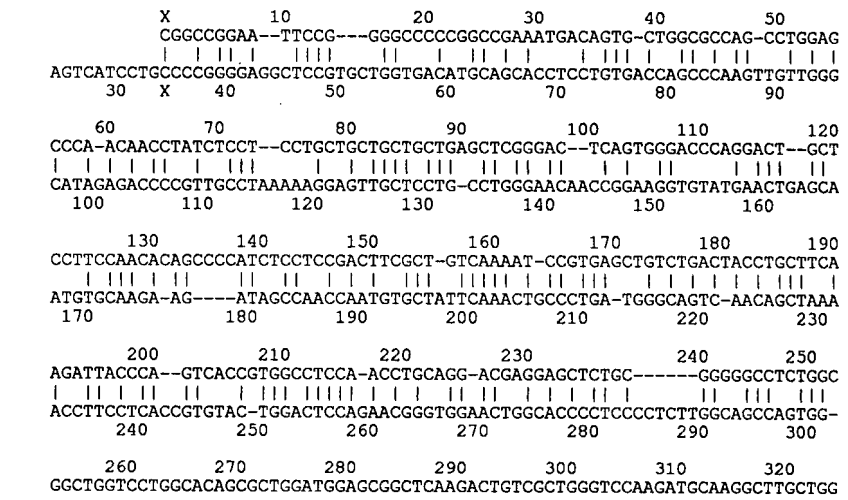


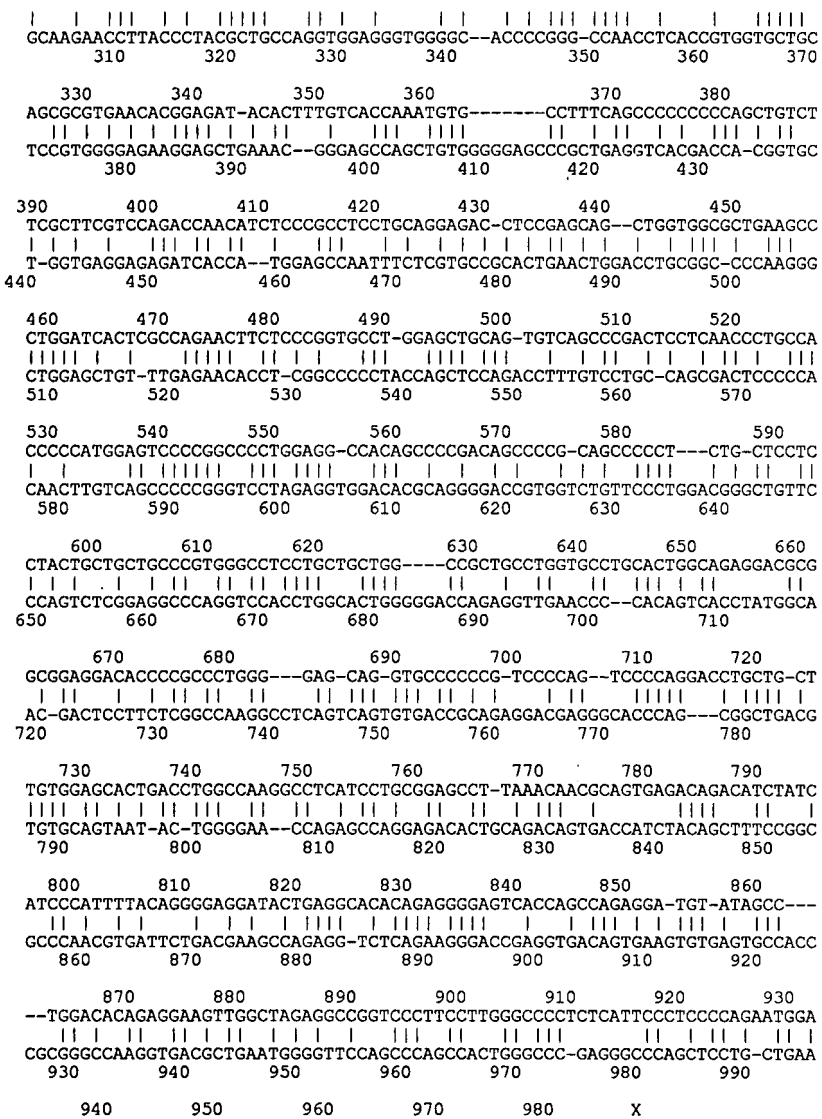


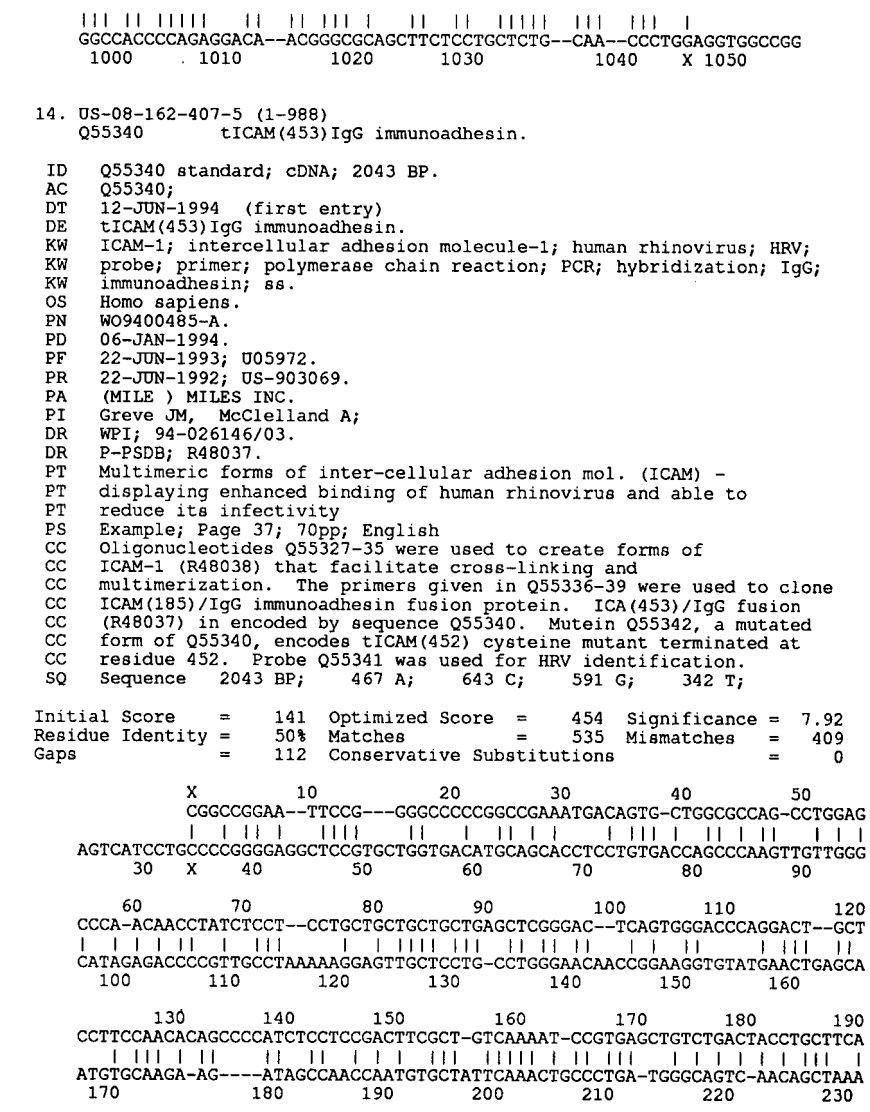
13. US-08-162-407-5 (1-988)
Q55342 tICAM(452) cysteine mutant.

ID Q55342 standard; cDNA; 1359 BP.
AC Q55342;
DT 12-JUN-1994 (first entry)
DE tICAM(452) cysteine mutant.
KW ICAM-1; intercellular adhesion molecule-1; human rhinovirus; HRV;
KW probe; primer; polymerase chain reaction; PCR; hybridizations; ss.
OS Homo sapiens.
PN WO9400485-A.
PD 06-JAN-1994.
PF 22-JUN-1993; U05972.
PR 22-JUN-1992; US-903069.
PA (MILE) MILES INC.
PI Greve JM, McClelland A;
DR WPI; 94-026146/03.
PT Multimeric forms of inter-cellular adhesion mol. (ICAM) -
PT displaying enhanced binding of human rhinovirus and able to
PT reduce its infectivity
PS Example; Page 47; 70pp; English
CC Oligonucleotides Q55327-35 were used to create forms of
CC ICAM-1 (R48038) that facilitate cross-linking and
CC multimerization. The primers given in Q55336-39 were used to clone
CC ICAM(185)/IgG immunoadhesin fusion protein. ICA(453)/IgG fusion
CC (R48037) in encoded by sequence Q55340. Mutein Q55342, a mutated
CC form of Q55340, encodes tICAM(452) cysteine mutant terminated at
CC residue 452. Probe Q55341 was used for HRV identification.
SQ Sequence 1359 BP; 294 A; 418 C; 411 G; 236 T;

Initial Score = 141 Optimized Score = 453 Significance = 7.92
Residue Identity = 50% Matches = 531 Mismatches = 415
Gaps = 108 Conservative Substitutions = 0



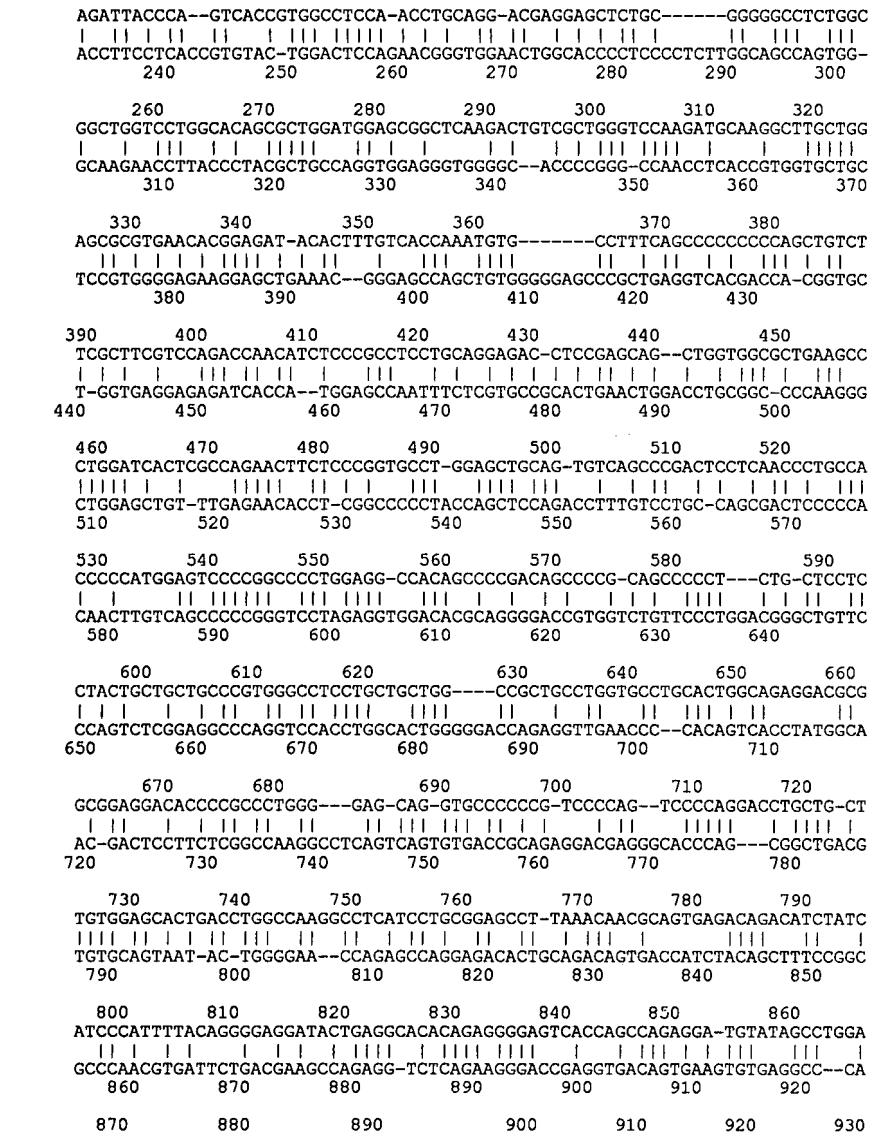


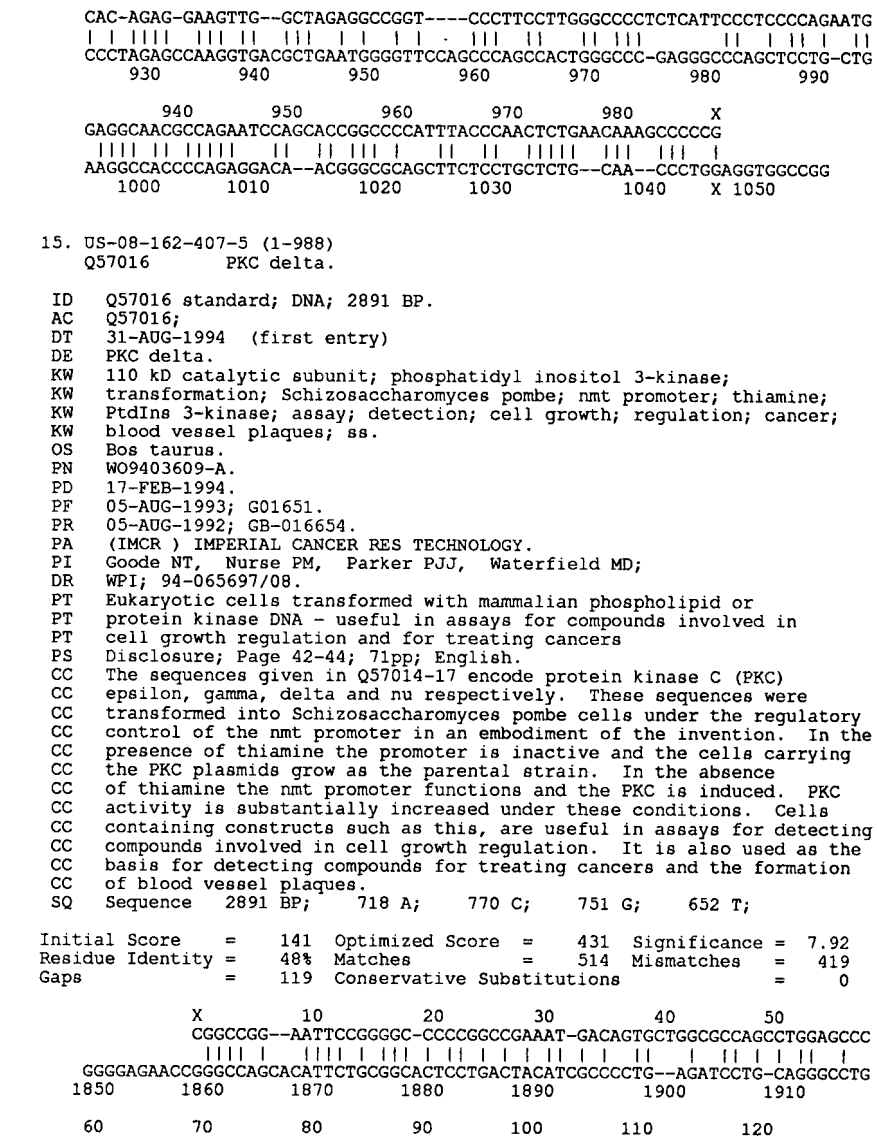


14. US-08-162-407-5 (1-988)
Q55340 tICAM(453)IgG immunoadhesin.

ID Q55340 standard; cDNA; 2043 BP.
AC Q55340;
DT 12-JUN-1994 (first entry)
DE tICAM(453)IgG immunoadhesin.
KW ICAM-1; intercellular adhesion molecule-1; human rhinovirus; HRV;
KW probe; primer; polymerase chain reaction; PCR; hybridization; IgG;
KW immunoadhesin; ss.
OS Homo sapiens.
PN WO9400485-A.
PD 06-JAN-1994.
PF 22-JUN-1993; U05972.
PR 22-JUN-1992; US-903069.
PA (MILE) MILES INC.
PI Greve JM, McClelland A;
DR WPI; 94-026146/03.
DR P-PSDB; R48037.
PT Multimeric forms of inter-cellular adhesion mol. (ICAM) -
PT displaying enhanced binding of human rhinovirus and able to
PT reduce its infectivity
PS Example; Page 37; 70pp; English
CC Oligonucleotides Q55327-35 were used to create forms of
CC ICAM-1 (R48038) that facilitate cross-linking and
CC multimerization. The primers given in Q55336-39 were used to clone
CC ICAM(185)/IgG immunoadhesin fusion protein. ICA(453)/IgG fusion
CC (R48037) in encoded by sequence Q55340. Mutein Q55342, a mutated
CC form of Q55340, encodes tICAM(452) cysteine mutant terminated at
CC residue 452. Probe Q55341 was used for HRV identification.
SQ Sequence 2043 BP; 467 A; 643 C; 591 G; 342 T;

Initial Score = 141 Optimized Score = 454 Significance = 7.92
Residue Identity = 50% Matches = 535 Mismatches = 409
Gaps = 112 Conservative Substitutions = 0

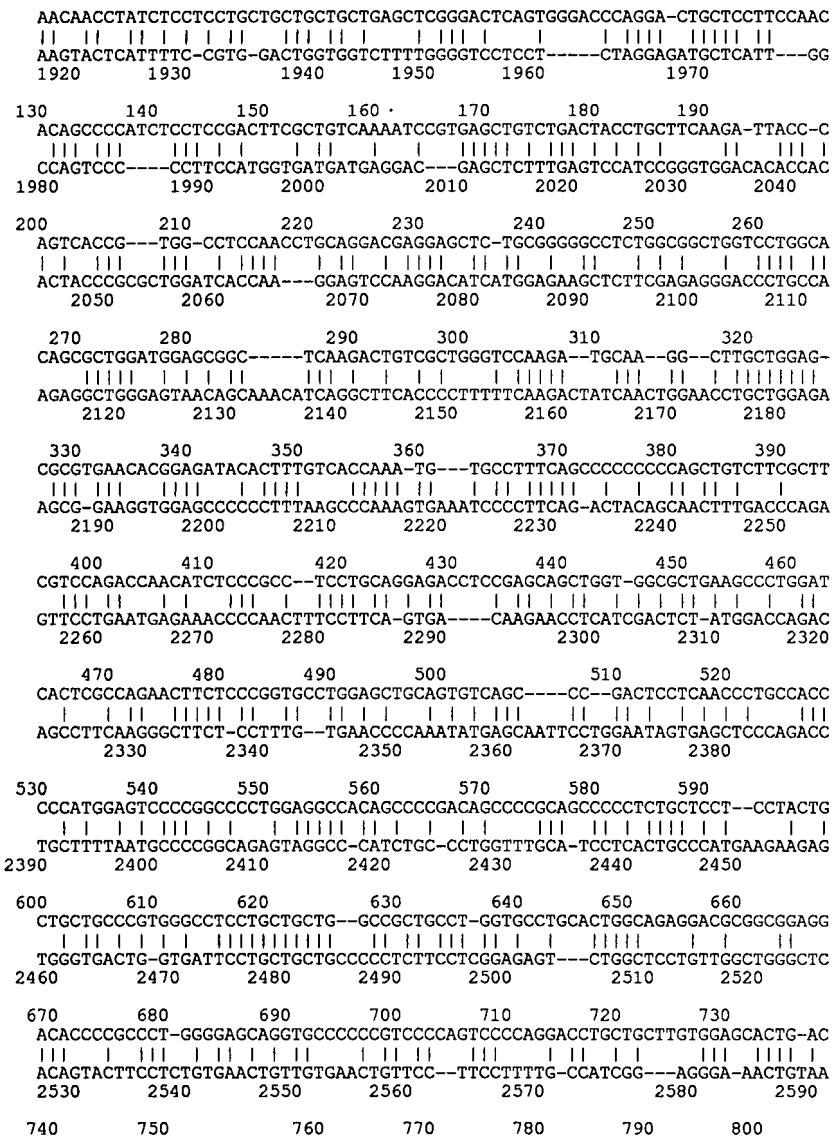


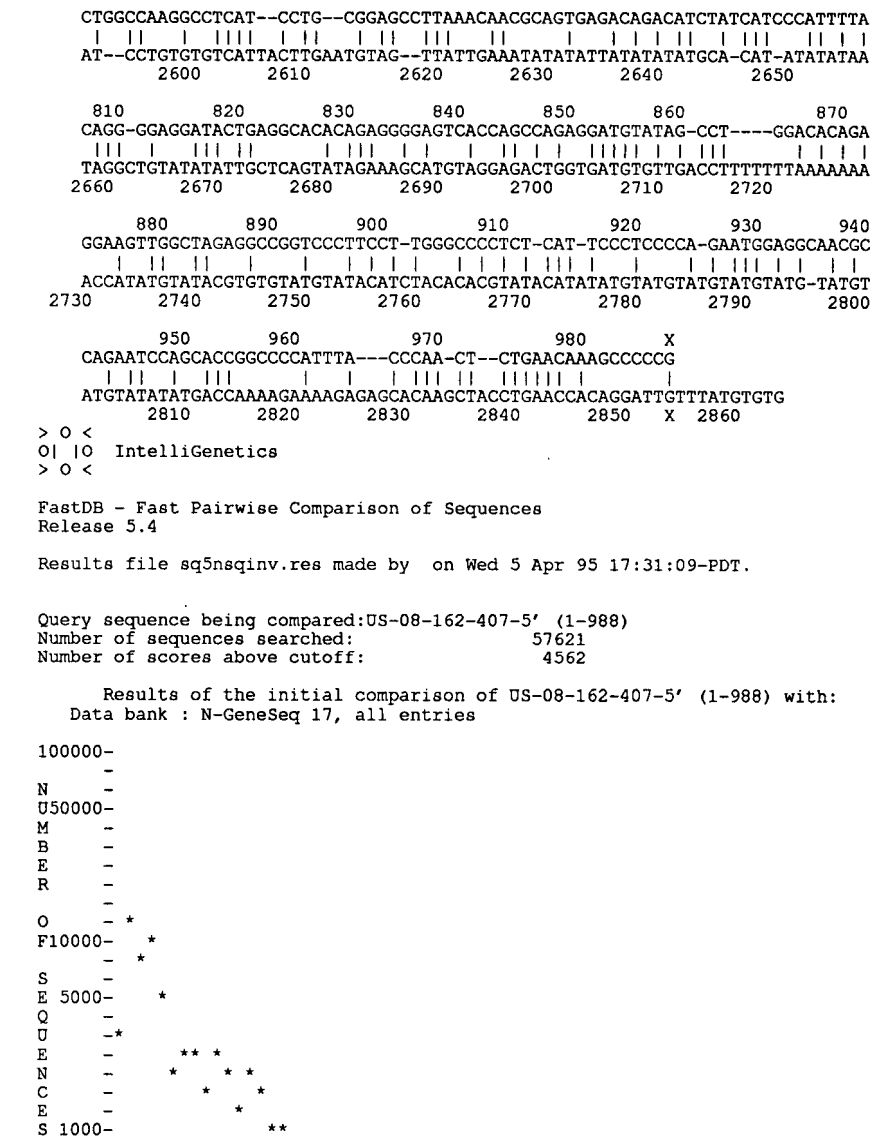


15. US-08-162-407-5 (1-988)
Q57016 PKC delta.

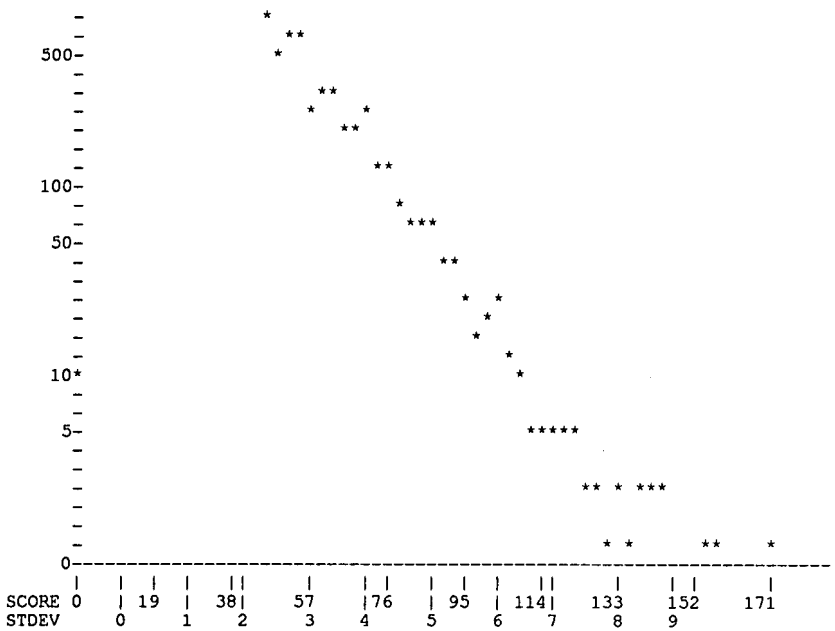
ID Q57016 standard; DNA; 2891 BP.
AC Q57016;
DT 31-AUG-1994 (first entry)
DE PKC delta.
KW 110 kD catalytic subunit; phosphatidyl inositol 3-kinase;
KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;
KW blood vessel plaques; ss.
OS Bos taurus.
PN WO9403609-A.
PD 17-FEB-1994.
PF 05-AUG-1993; G01651.
PR 05-AUG-1992; GB-016654.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PI Goode NT, Nurse PM, Parker PJJ, Waterfield MD;
DR WPI; 94-065697/08.
PT Eukaryotic cells transformed with mammalian phospholipid or
PT protein kinase DNA - useful in assays for compounds involved in
PT cell growth regulation and for treating cancers
PS Disclosure; Page 42-44; 71pp; English.
CC The sequences given in Q57014-17 encode protein kinase C (PKC)
CC epsilon, gamma, delta and nu respectively. These sequences were
CC transformed into Schizosaccharomyces pombe cells under the regulatory
CC control of the nmt promoter in an embodiment of the invention. In the
CC presence of thiamine the promoter is inactive and the cells carrying
CC the PKC plasmids grow as the parental strain. In the absence
CC of thiamine the nmt promoter functions and the PKC is induced. PKC
CC activity is substantially increased under these conditions. Cells
CC containing constructs such as this, are useful in assays for detecting
CC compounds involved in cell growth regulation. It is also used as the
CC basis for detecting compounds for treating cancers and the formation
CC of blood vessel plaques.
SQ Sequence 2891 BP; 718 A; 770 C; 751 G; 652 T;

Initial Score = 141 Optimized Score = 431 Significance = 7.92
Residue Identity = 48% Matches = 514 Mismatches = 419
Gaps = 119 Conservative Substitutions = 0





```
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq5nsqinv.res made by  on Wed 5 Apr 95 17:31:09-PDT.


Query sequence being compared:US-08-162-407-5' (1-988)
Number of sequences searched:                 57621
Number of scores above cutoff:                 4562

     Results of the initial comparison of US-08-162-407-5' (1-988) with:
   Data bank : N-GeneSeq 17, all entries

100000-
      -
N     -
U50000-
M     -
B     -
E     -
R     -
      -
O     - *
F10000-   *
      -  *
S     -
E 5000-    *
Q     -
U     -*
E     -      ** *
N     -     *    * *
C     -         *    *
E     -              *
S 1000-                **
```



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	2		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	18	14	14.81

Times:	CPU	Total Elapsed
	00:04:06.02	00:04:07.00

Number of residues: 24347505
Number of sequences searched: 57621
Number of scores above cutoff: 4562

Cut-off raised to 11.

Cut-off raised to 17.
Cut-off raised to 26.
Cut-off raised to 31.
Cut-off raised to 36.
Cut-off raised to 41.
Cut-off raised to 45.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 10 standard deviations above mean ****					
1. Q23092	Antigen tc-7a gene.	543	171	244	10.33	0
2. Q51731	Plasmid pCisEBON for subcloni	10596	170	454	10.26	0
	**** 9 standard deviations above mean ****					
3. Q14184	N.clavipes dragline silk prot	1995	157	449	9.39	0
4. N92629	DNA fragment containing acyB	2729	153	457	9.12	0
	**** 8 standard deviations above mean ****					
5. Q29269	Human calcium channel 27980/1	6232	144	442	8.51	0
6. Q37817	Sequence encoding the alpha 1	7362	143	442	8.44	0
7. Q54631	Human oncogene bcl-2 coding s	5086	141	321	8.31	0
8. N81292	Sequence of bcl-2 cDNA corres	5105	140	320	8.24	0
9. Q05884	Plasmid clone 409.2.	633	137	238	8.03	0
10. Q05883	Plasmid clone pUC-SK1.	6767	137	381	8.03	0
	**** 7 standard deviations above mean ****					
11. Q03842	LFA-1 alpha subunit gene	5139	134	378	7.83	0
12. Q10388	Encodes skin-specific murine	1912	132	431	7.70	0
13. Q20541	Mouse retinoic acid receptor	2740	132	417	7.70	0
14. Q14498	Clone 1 for truncated poly Ig	1876	130	263	7.56	0
15. Q05272	Sequence encoding secreted GP	2183	126	368	7.29	0
16. Q62199	Human platelet GPIIIa surface	3170	126	429	7.29	0
17. N80476	Human Vpre-B gene.	503	124	229	7.16	0
18. N80470	Gastric inhibitory polypeptid	524	124	237	7.16	0
19. Q05686	Islets of Langerhans cell clo	3243	122	412	7.02	0
20. Q31188	Alpha 6A integrin subunit cDN	5629	122	184	7.02	0
	**** 6 standard deviations above mean ****					
21. Q56906	DNA encoding a glycosyltransf	2043	121	436	6.95	0
22. Q13330	Glycosyltransferase gene in c	2043	121	436	6.95	0
23. N91779	Sequence encoding mouse lamin	9520	121	278	6.95	0
24. Q04735	Gene fragment encoding human	1226	120	232	6.89	0
25. N91772	Human androgen receptor cDNA.	3569	120	446	6.89	0
26. Q54431	Barley lectin cDNA clone BLc3	972	119	345	6.82	0
27. Q27764	Adrenodoxin reductase gene (p	5089	118	440	6.75	0
28. Q67602	Retinoblastoma protein Ap10.	4868	116	407	6.62	0
29. Q51557	Loricrin gene.	6530	116	449	6.62	0
30. Q40969	Rabbit skeletal muscle ADP-ri	1140	115	442	6.55	0
31. Q25975	MH mutant porcine ryanodine r	15377	115	439	6.55	0
32. N71006	Sequence encoding the N-termi	1365	113	343	6.41	0
33. N71008	Sequence encoding fusion prot	1389	113	367	6.41	0
34. Q13674	E2A.E12 transcript.	3518	113	292	6.41	0
35. Q04549	Type I interferon receptor co	3888	113	441	6.41	0
36. Q47380	Sequence which corresp. to L-	12752	113	429	6.41	0
37. N71005	Sequence encoding the N-termi	959	112	366	6.35	0

38.	N60426	Sequence encoding bovine inhi	1182	112	318	6.35	0
39.	Q04081	Vector pASPcg-SV(10)	2720	111	440	6.28	0
40.	N90338	Sequence of human muscular dy	12923	111	393	6.28	0
41.	Q54841	Sequence encoding inner nucle	6306	110	450	6.21	0
42.	Q20733	pAD-CMV1 expression vector.	6414	109	441	6.14	0
43.	Q50991	Human prohibitin gene.	1020	108	284	6.08	0
44.	Q25264	C.glutamicum-derived temp-sen	2958	108	428	6.08	0
45.	Q53993	Myotonic dystrophy (DM) gene	3407	108	457	6.08	0

1. US-08-162-407-5' (1-988)
   Q23092 Antigen tc-7a gene.

ID Q23092 standard; DNA; 543 BP.
AC Q23092;
DT 17-AUG-1992 (first entry)
DE Antigen tc-7a gene.
KW Oocysts; MAb 12-07; sporozoite; ss.
OS Eimeria tenella.
PN WO9204460-A.
PD 19-MAR-1992.
PF 05-SEP-1991; U06430.
PR 12-SEP-1990; US-581693.
PA (GENE-) GENEX CORP.
PI Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;
PI Ruff MD, Augustine PC, Danforth HD;
DR WPI; 92-114365/14.
DR P-PSDB; R22392.
PT Vaccine against avian coccidiosis - comprising recombinant
PT Eimeria antigen ac-1b or ac-6b gene, or microorganisms expressing
PT them
PS Claim 5; Page 39 + Fig 5; 56pp; English.
CC To identify antigens of E. tenella, expression libraries were
CC prepd. in lambda vector, lambda gt11, using cDNA prepd. from polyA
CC mRNA isolated from E. tenella oocysts. The cDNA expression
CC library was screened with monoclonal antibody (MAb) 12-07 which was
CC raised against the sporozoite stage of E. tenella. The library
CC to be screened was plated on a host that allows lysis and plaque
CC formation. During induction of the antigens encoded by the phage,
CC the plaques were identified by screening the filters with Mab 12-07.
CC The cDNA inserts from the MAb 12-07 positive phage were cloned into
CC bacteriophage M13 and subjected to sequence analysis. Following
CC sequence analysis, E. tenella antigen tc-7a was identified.
SQ Sequence 543 BP; 229 A; 135 C; 161 G; 18 T;

Initial Score = 171 Optimized Score = 244 Significance = 10.33
Residue Identity = 49% Matches = 285 Mismatches = 232
Gaps = 54 Conservative Substitutions = 0

```
      140      X 150       160       170       180       190       200
    TCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATGGGATGATAGATGTCTG-T
              | ||          | |    || |  |       | | ||   || || ||| |   |
              AGAGAAGAAGACAAAAGAGAAGAAGAAGACAAAAGAGAAGAA---GAAGA-AGAAGAAAGAG
              X       10        20        30        40            50

     210       220       230       240       250       260           270
    CTCACTGCGTTGTTTAAGGCTCCGCAGGATGAGGCCTTG-GCCAGGTCAGTGCTCCA-CA--AGCAGCAGGT
       |  | |  |   |||       || |  ||     | | |||  |  |||  || ||  ||||||||
    AAGAGAGAGGAGAAGAAGAAGAAAAAGAAGAAGAAAGAGCGGCAGCAC-CTGCAGCAGCAACAGCAGCAGCA
    60        70        80        90       100       110       120
```



```
       280       290       300       310       320       330       340
    CCTGGGGACTGGGGACGG--GGGGCACCTGCTCC--CCAGGGCGGGGTGTCCTCCGCCGCGTCCTCTGCCAG
    ||||  | | |   || |  |  ||| | ||  |  | | |  ||   |   |  | ||  |  |  ||||
    CCTGCAG-CAG-CAACAGCAGCAGCAACAGCAGCTGCAACAGC--AGCAGCTACAGCAACACCAGCTGCAGC
130         140       150       160       170       180       190

      350       360       370       380       390       400       410
    TGCAGGCACCAG--GCAGCGG--CCAGCAGCAGGAGGCCCACGGGCAGCAGCAGTAGGAGGAGCAGAGGGGG
    |||| |||||||  |||||  |  ||||| |||| ||    ||   ||||||| ||| ||  | |||||     |
    TGCA-GCACCAGCTGCAGCAGCACCAGCTGCAGCAG---CAGCAGCAGCAACAGGAGCTGCAGCAGCACCAG
    200       210       220       230       240       250       260

          420       430       440       450       460       470       480
    CTGC-GGGGCTGTCGGGGCTGTGGCCTCCAGGGGCCGGGGACTCCATGGGGGTGGCAGGGTTGA-GGAGTCG
    |||| | | | | ||  || | |  |  |||| || |       | | ||  | || ||      |  ||  |
    CTGCAGCGACAGCCGCAGCAG-GAGCAGCAGGAGCAG------CAAAGGCAGCGGAAGCACGAACAGAAACC
      270       280       290       300             310       320       330

          490       500       510       520       530       540
    GGCTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATCCAG-GGCTTCAGC-GC--CACCAG
    |   ||  | ||||   || ||| | | || ||     ||||     | ||  ||  |||| || ||  || |||
    GAAGGA-GCAGCAGGAGCA-GCAGCAGAAGCAGAGAAAGCGAAAACACAAGCAGCAACAGCAGCAACAACAG
          340       350       360       370       380       390       400

    550       560       570       580       590       600       610
    CTGCT-CGG--AGGTCTCCTGCAGGAGGCGGGAG-ATGTTGGTCTGGACGAAGCGAAGACAGC-TGGGGGGG
    | ||| | |  ||  |  | |||| |  | | || | ||  | ||||  ||   | || |||| |     |
    CAGCTGCAGCAAGAGCAGCAGCAGCA-TCAGCAGCAGGTAAGCCTGGGGGACATGCAG-CAGCATCAACAGC
          410       420       430       440       450       460       470

      620       630       640       650       660       670       680
    GGCTGAAAGGCACATTTGGTGACAAAGTGTATCTCCGTGTTCACGC-GCTCCAGCAAGCCTTGCATC-TTGG
     ||  ||| |||     |  || |||| |          |   |||| ||  |||||     |||||  ||
    AGC--AAAAGCAGCGCAG--GAAAAAGCGGCGAAAGCAGCAAACGCAGCAACAGCAGCAACAGCATCAAAGG
            480       490       500       510       520       530

      690       700
    ACCCAGCGACAGTCT
    |   |
    AATAA
    540  X
```

2. US-08-162-407-5' (1-988)
   Q51731 Plasmid pCisEBON for subcloning huHGF variants.

ID Q51731 standard; DNA; 10596 BP.
AC Q51731;
DT 31-MAY-1994 (first entry)
DE Plasmid pCisEBON for subcloning huHGF variants.
KW Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW proteolysis resistant; liver; malignancy; CMV-driven;
KW Cytomegalovirus; episomal expression plasmid; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT enhancer 1..611
FT /*tag= a
FT /note= "CMV enhancer/promoter"
FT promoter 758..775

```
FT   /*tag=  b
FT   /label= SP6_promoter
FT   misc_feature    845..849
FT   /*tag=  c
FT   /note= "SP6 RNA start"
FT   misc_feature    902..966
FT   /*tag=  d
FT   /function= cloning_linker
FT   polyA_signal    967..1107
FT   /*tag=  e
FT   /note= "SV40 poly A"
FT   misc_feature    1108..1531
FT   /*tag=  f
FT   /function= SV40_origin
FT   misc_feature    1580..4189
FT   /*tag=  g
FT   /label= EBNA-1
FT   misc_feature    4190..6374
FT   /*tag=  h
FT   /function= oriP
FT   repeat_region   4295..4887
FT   /*tag=  i
FT   /note= "family of repeats"
FT   misc_structure  5866..5978
FT   /*tag=  j
FT   /note= "dyad region"
FT   terminator      6375..6457
FT   /*tag=  k
FT   /label= HSV_TK_terminator_3'-end
FT   CDS             6975..7975
FT   /*tag=  l
FT   /phenotype= neomycin_resistance
FT   /note= "Tn5 neomycin phosphotransferase gene"
FT   promoter        7975..8112
FT   /*tag=  m
FT   /label= TK_promoter
FT   misc_feature    8114..8594
FT   /*tag=  n
FT   /function= M13_ori
FT   misc_RNA        8595..10414
FT   /*tag=  o
FT   /label= delta_2a
PN   WO9323541-A.
PD   25-NOV-1993.
PF   17-MAY-1993; U04648.
PR   18-MAY-1992; US-884811.
PR   18-MAY-1992; US-885971.
PA   (GETH ) GENENTECH INC.
PI   Godowski PJ,  Lokker NA,  Mark MR;
DR   WPI; 93-386573/48.
PT   Hepatocyte growth factor variants - are resistant to proteolytic
PT   cleavage into its two-chain form, used to treat malignancies
PT   associated with HGF receptor
PS   Example 1; Fig 6; 87pp; English.
CC   Plasmid pCisEBON (a pRK5 derivative) is an episomal CMV driven
CC   expression plasmid. HuHGF variants with enhanced receptor binding
CC   activity were produced by site-directed mutagenesis. Stable
CC   populations of preferred HGF variants were obtained by transfecting
CC   human embryonic kidney 293 cells and then these were subcloned in
CC   pCisEBON. See R52940-R52949 for examples of pref. HGF variants.
```



```
 SQ   Sequence   10596 BP;    2625 A;    2571 C;    3024 G;    2376 T;

Initial Score    =    170  Optimized Score  =    454  Significance = 10.26
Residue Identity =    50%  Matches          =    529  Mismatches   =   418
Gaps             =    105  Conservative Substitutions              =     0

            X       10          20        30        40        50
            CGGGGGCTTTG---TTCAGAGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTG-CCTC
            ||| |||  ||    ||| ||  | |    ||    |  ||   ||    || |  |    |
   CAGAAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGAC
       2040      2050      2060      2070      2080      2090      2100

   60        70          80        90          100       110       120
   CATTCTGGGGAGG-GAATGA-GAGGGGCCCAAGGAAG----GGACCGGCCTCTAGCCAACTTCCTCTGTGTC
    |    | |||||  | | ||  ||| ||||  ||  |    ||| | |   | || ||       |  |||||
   GAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGAT-GGTGTC
       2110      2120      2130      2140      2150      2160      2170

      130       140       150       160       170       180       190
   CAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATG-GGATGA
   | |   | ||  ||      |     | |  | | ||  || |  ||   ||  |  || || | | | | ||
   CGG---AGAC-CCCAAAAACGTCCAAGTTGCAT-TGGCTG-CAAAGGGACCCACGGTGGAACAGGAGCAGGA
       2180      2190      2200      2210      2220      2230

      200       210       220       230       240       250       260
   -TAGATGTCTGTCTCACTGCGTTGTTTAAGGCTC---CGCAGGATGA-GGCCTTGGCCAGG-TCAG-TGCTC
     ||  |   |     |   | |   |||  |    |||||| || || | |  ||||   ||  ||
   GCAGGAGCGGGAGGGGC--AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAG
2240      2250      2260      2270      2280      2290      2300

         270       280       290       300       310       320
   CACAAGCAGCAGGTCCTG--GGGACTGG----GGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCG
    |   |||| |||  | |  ||| | ||    ||| | ||||||   |     ||| | || | | |    |
   GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAG-----CAG-GAGGAGGGGCAGGAG
2310      2320      2330      2340      2350      2360      2370

    330       340       350       360       370       380       390
   CCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGCCC---ACGGGCAGCAGCAGTAG
     ||       ||  | |||||  |  |||  || |  |||||  ||||||  |   | ||||||  |||||  ||
   GGGCAGGAGGGGCAGGAGCAGG-AGGAGG-GGCAG-GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG
    2380      2390      2400      2410      2420      2430      2440

      400       410       420       430       440       450       460
   GAGGAGCA-GAGGGGGCTGCGGGGCTGTCG-GGGCTGTGGCCTCCAGGGGCCGG-GGA-CTCCATG-GGGGT
   |||| ||| ||||||   | |||||  |  |  ||  | ||    |||| || || |||    || | ||||
   GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGG--GGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
     2450      2460      2470      2480      2490      2500      2510

         470       480       490       500       510       520
   GGCAG--GGTTGAGG---AGTCGGG-CTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATC
    | ||  ||  ||||   ||  ||| | |     ||||   ||||    |||  | |  | || | | |
   AGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
     2520      2530      2540      2550      2560      2570      2580

  530       540       550       560       570       580       590       600
   CAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGA
     ||||   || | ||  ||| |   |  || |    |||||||| |||  | |  || | ||| |    |
   AGGGGCAGGAGGGGCAGGAGCAG---GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG
   2590      2600      2610      2620      2630      2640      2650
```

```
         610         620         630         640         650         660
AGACAGCTGGGG-GGGGGC-TGAAAGGCA----CATTTGGTGACAAAGTGTATCTCCGTGTTCACGCGCTCC
|| |||  ||||  || ||  ||  ||||    ||   || | ||  | | |      |      | | ||
AG-CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGG-GGCA-GGAGCAGGAGGGGCAGGAGGGGCAGG
  2660       2670       2680       2690        2700       2710       2720

 670        680         690         700         710         720        730
AGCAAGCCTTGCATCTTGGAC-CCAGC--GACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGA---CC
||||  |    |||     || |   |||  ||  | |  |||  ||   | |  | |  | | |||||    |
AGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGC
  2730       2740       2750       2760       2770       2780       2790

      740        750         760         770         780          790
AGCCGCCAGAGGCCCCCGCA-GAGCTCCTCGTCCTGCAGGTTGGAGGCCACGGTGA-----CTGGGTAATCTT
||  ||  ||||      ||| |||   |  |  | | |||  | |||  | ||  |     | ||      |
AGGAGCAGGAGG-AGGGGCAGGAGGGGCAGGAGCAGGAGG-GGCAGG--AGGGGCAGGAGCAGGAGGGGCAG
2800       2810       2820       2830       2840        2850       2860

800        810         820        830         840        850        860
  GAAGCAGGTAG--TCAGACAGCTCACGGATTTTGACA-GCGAAGTCGGAGGAGATGGGGCTGTGTTGGAAGG
  || |  || ||  |||   |  || |||     | || | | ||  ||||| |  || || |    || |||
  GAGG--GGCAGGAGCAGGAGGGGCA-GGA--GGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGG
      2870       2880       2890        2900       2910       2920       2930

 870        880         890        900         910        920        930
  AGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGA-TAGGTTGTTGGGCTCCAG
  |||||     ||   | | | ||  ||||    ||||  |  || || |  ||  |||| | |||    || |
  AGCAGGAGGTGGAGGC-CGGGGT--CGAG-GAGGCAGTGGAGGCCGGGGTCGAGGAGGTAG-TGGAGGCCGG
        2940       2950        2960       2970       2980       2990

940        950         960           970         980       X
  GCTGGCGCCAGCACTGTCATTTCGGCCGGG----GGCCCCGGAA-TTCCGGCCG
  | | | |   | | ||  |   || | |||   || |  | ||   || | |
  GGTCGAGGAGGTAGTG-GAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGT
3000       3010        3020       3030       3040       3050       3060
```

3. US-08-162-407-5' (1-988)
   Q14184 N.clavipes dragline silk protein-2 coding sequence

```
ID   Q14184 standard; DNA; 1995 BP.
AC   Q14184;
DT   15-JAN-1992  (first entry)
DE   N.clavipes dragline silk protein-2 coding sequence.
KW   protein superfibre; major ampullate silk; orb web spider; ss.
OS   Nephilia clavipes.
FH   Key             Location/Qualifiers
FT   CDS             1..1788
FT   /*tag=  a
FT   /product= silk protein-2
PN   EP-452925-A.
PD   23-OCT-1991.
PF   18-APR-1991; 106217.
PR   20-APR-1990; US-511792.
PA   (UYWY-) UNIV OF WYOMING.
PI   Lewis RV, Xu M, Hinman M;
DR   WPI; 91-312199/43.
DR   P-PSDB; R14309.
PT   DNA encoding spider silk protein-1 and 2 and variants - isolated
PT   from Nephila clavipes, for prodn. of spider silk protein and
```



```
PT   fibres having desired characteristics
PS   Claim 5; Page 30; 48pp; English.
CC   A N.clavipes major ampullate gland cDNA library was screened with
CC   a degenerate probe based on the pentapeptide GYGPG. The largest
CC   positive clone (p6B) was subcloned and sequenced. Plasmid pMB2 is
CC   a pBluescriptSK+ plasmid with a spider silk 2 insert 173 bp
CC   shorter than in p6B. E.coli SURE cells containing plasmid pMB2 were
CC   deposited as ATCC 68568.
CC   See also Q14183 and Q14185.
SQ   Sequence   1995 BP;    563 A;    518 C;    586 G;    328 T;
```

```
Initial Score    =    157  Optimized Score  =    449  Significance =  9.39
Residue Identity =    51%  Matches          =    544  Mismatches   =   387
Gaps             =    133  Conservative Substitutions               =     0

            X         10          20         30         40         50
            CGGGGGC--TTTG-TTCAG--AGTTGGGTAAATGGGGCCGGTGCTGGATTCTGGCGTTGCCT
              |  || |  | ||   |||  |||   | |   |  ||||  || |||  ||||    |
   ACCAGGACAACAAGGACCATCTGGAGCAGGCAGTGCAGCAGCAGCAGCCGCAGCAGGA-CCTGG----ACAA
    900        910        920        930        940        950          960

    60         70         80         90         100        110        120
   CCATTCTGGGGAGGGAATGAGAGGGGCCCAAGGAAGGGACCG--GCCTCTAGCCAACTTCCTCTGTGTCCAG
   | |   |  |||||  ||| ||   |  |||  ||||  | |  |  | | | | |   |  |   ||||||
   CAAGGATTAGGAGGTTATG-GACCAGGACAA-CAAGGTCCAGGAGGATATGGACCAGGACAACAAGGTCCAG
       970        980        990        1000       1010       1020       1030

   130        140        150        160        170        180        190
   GCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTA-AAATGGGATGATAG
   |    | ||   |  | | |||  ||  |      ||| |||   |    |||| | ||   |||  |  |
   G-AGGATATGGACCAGGTAGTGCATCTGCAGCAGCAGCCGCAG---CAGGACCTGGACAACAAGGACCA-GG
       1040       1050       1060       1070       1080       1090       1100

  200        210        220        230        240        250        260
   ATGTCTGTCTCACTGCGTTGTTTAAGGCTCCGCAGGATGAGGCCTTG-GCCAG-GTCAG-TGCTCCACAAGC
   | |  | |   ||  |      ||||  |  | |||  ||||  ||   | |   |||  ||  ||   ||
   A-GGATATGGACCT--GGACAACAAGGACCATCTGGACCAGGCAGTGCATCTGCAGCAGCAGCCGCAGCCGC
         1110       1120       1130       1140       1150       1160

   270        280        290        300        310        320        330
   AGCAGGTCCTGGGGACTGGGGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCC--GCC-GC---GTC
   |||||| || || |  |  ||||  ||   ||  | |  |||||   || | | | ||  | |  |   | |
   AGCAGGACCAGGAGGATATGGACCAGGACAACAAGGT--CCAGG--AGGATATGCACCAGGACAACAAGGAC
1170       1180       1190       1200       1210       1220       1230

         340        350        360        370        380        390
   C-TCTG--CCAGTGCAG-GCACCAG--GCAGCGGC--CAGCAGCAGGAGGCCCACGGGCA---GCAGCAG--
   | ||||  |||| |||| ||| | |  |||||  ||  |||| |||| ||| ||| | | |   | | |||
   CATCTGGACCAG-GCAGTGCATCTGCAGCAGCAGCCGCAGCCGCAGGACCAGGAGGATATGGACCAGGA
  1240       1250       1260       1270       1280       1290       1300

        400        410        420        430        440        450
  -TAGGAGGAGCA-GAGG--GGGCTGCGGGGCTGTCGGGGC--TGTGGCCTCCAGG--GGCCGGGGACTCCAT
    |  |||  || ||||    ||  | || |      || |  | |||   |||||  |  | |   |  ||
   CAACAAGGTCCAGGAGGATATGCACCAGGACAACAAGGACCATCTGG--ACCAGGCAGTGCAGCAGCAGCA-
 1310       1320       1330       1340       1350       1360       1370

   460        470        480        490        500        510        520
   GGGGGTGGCAGGGTTGAGGAGTCGGGCTGACACTGCAGCTCCAG-GCACCGGGAGAAGTTCTGG-CGAGTGA
   |  | || ||  ||  ||||  | ||  || | || |   |||| ||| | || ||   ||||| |  |  |
```

```
    GCAGCTGCCA--GTGCAGGA-CCTGGTGGATA-TGGA---CCAGCGCAACAGG-GACCATCTGGTCCTGGAA
  1380        1390        1400        1410         1420        1430       1440

     530        540        550         560          570        580        590
    TC-CAGGGCTTCAGCGCCACCAGCTG--CTCGGAGGT-----CTCCTGCAGGAGGCGGGAGATGTTGGTCTGG
    || ||  |||||||||  |  |||| |  |  ||||||       || |||  |   ||   |  | | |||
    TCGCA--GCTTCAGCTGCTTCAGCAGGACCTGGAGGTTATGGACCAGCACAACAAGGACCAGCTGGATATGG
          1450       1460       1470       1480       1490       1500       1510

             600          610        620        630        640        650
    AC---GAAGCGAAG--ACAG-CTGGGGGGGGGC-TGAAAGGCACATTTGGTGACAAAGTGTATCTCCGTGTT
      |   ||||||  ||   |||  ||  |  ||  ||  ||   |||  ||  | | | | ||   ||||    ||
    GCCTGGAAGCGCAGTAGCAGCCTCTGCCGGTGCAGGATCTGCAGGTTATG-GGCCAGGT---TCTCAAGCTT
         1520       1530       1540       1550       1560          1570

     660        670        680        690        700        710        720
    CACGC-GCTCCAGCAAGCCTTGCATCTTGGACCCAGCGACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCA
    | |||  ||   |  |  | ||  ||  |||     ||||   |||  |    ||  | | |||  |||   ||||  |
    C-CGCTGCAGCTTCTCGTCTGGCTTCT------CCAGATTCAGGCGCTAG-AGTTGCAT-CAG--CTGTTTCT
1580        1590       1600            1610       1620       1630       1640

    730        740        750        760        770        780        790
    GGACCAGCCGCCAGAGGCCCCCGCAGAGCTCCT-CGTCCTGCAG-GTTGGAGGCCACGGTGACTGGGTAATC
          |   ||||  |||||     ||  || || | |  |  || |||     |  |||  ||  ||  | |
    AACTTGGTATCCAGTGGCCCAACTAGCTCTGCTGCCTTATCAAGTGTTATCAGTAACGCTG--TGTCTCAAA
         1650       1660       1670       1680       1690       1700       1710

    800        810        820        830        840        850        860
    TTGAAGCAGGTAGT-CAGACAGCTCACG----GATTTTGACAGCGAAG-TCGGAGGAGAT-GGGGCTG-TGT
    |||  ||| ||| | | |    |||  |   ||| |   ||   ||| ||   ||| || |   ||| |
    TTGGCGCAAGTAATCCTGGTCTCTCTGGTTGCGATGTCCTCATTCAAGCTCTCTGGAAATCGTTTCTGCTTG
         1720       1730       1740       1750       1760       1770       1780

          870        880         890        900        910        920
    TGGAAGGAGCAGTCCTGGGTCC--CACTGAGTCCCG--AGCTCAGCAGCAGC--AGCAGGAGGAGATAGGTT
    || ||  | |  | ||   |||  || || |||  |  |  |  | ||| ||    |||   |    | |||
    TGTAACCATCCTTTCTTCATCCAGCATTG-GTCAAGTTAATTATGGAGCGGCTTCTCAGTTCG-CCCAAGTT
       1790       1800       1810       1820       1830       1840       1850

    930        940         950        960        970        980          X
    GTTGGGCTCCAGGCTG--GCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGGAAT-TCC--GGCCG
    || ||   |||  |||     |   ||| | | |||   |              ||| | |  ||
    GTCGG---CCAATCTGTTTTGAGTGCATTTTAATT---GAAAAATTTATTAAAATATGCATGGATTTTCTAG
          1860       1870       1880         1890       1900       1910  X

    CCTG
1920
```

4. US-08-162-407-5' (1-988)
   N92629 DNA fragment containing acyB gene.

ID N92629 standard; DNA; 2729 BP.
AC N92629;
DT 15-MAY-1990 (first entry)
DE DNA fragment containing acyB gene.
KW Macrolide; antibiotics; Streptomyces; acylation; ds.
OS Streptomyces sp.
PN EP-345546-A.
PD 13-DEC-1989.

PF 24-MAY-1989; 109425.
PR 24-MAY-1988; JP-125091.
PR 03-MAR-1989; JP-050120.
PA (SAOC) Sanraku Inc.
PI Arisawa A, Kawamura N, Kojima I, Okumura Y, Okamura K, Tone H,
PI Okamoto R;
DR WPI; 89-365353/50
PT DNA coding for enzyme acylating 4"-position of macrolide antibiotics -
PT contg. acyB gene and derived from Streptomyces microorganism.
PS Claim 7; Page 28-33; 54pp; English.
CC DNA fragment contains two open reading frames, one upstream, one down-
CC stream, coding for acyB1 and acyB2 genes respectively.
SQ Sequence 2729 BP; 475 A; 891 C; 962 G; 401 T;

Initial Score = 153 Optimized Score = 457 Significance = 9.12
Residue Identity = 51% Matches = 551 Mismatches = 373
Gaps = 137 Conservative Substitutions = 0

```
             X       10         20         30         40         50
             CGGGGGCTTTGTTCA-GAGTTGGGTAAATGGGGCCGGTGCTG----GATTCTGGCGTTGCCT
             | |  ||   | |||    |   ||| ||  | ||||| |  |    | ||| |||   ||
    CCGAGGCGAGCAGCAGC-GCGATCACCGGGACGGTCAA-CGCGCCGGGGTCGTAGTGGTTC-GGCACCGCGA
       400   X  410        420        430        440        450        460

     60         70         80         90        100        110        120
    CCAT-TCTGGGGAGGGAA-TGAGAGGGGCCCAAGGAAGG-GACCGGCCTCTAGCCAACTTCCTCTGTGTCCA
     | |  | | | |      |||||   || | | | |||  |  || | |  |||  | |||   ||  ||
    ACGTCACCGCGAACACCGCTGAGAACAGCGCGACGGCAGGCCAGGGGACGCGGGCC--CCTCCAGCGT-CCCC
       470        480        490        500        510        520        530

     130        140        150        160        170        180        190
    GGCTATACATCCTCTGG--CTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCC---CTGTAAAATG-GGA
    || |   ||   |||||  |  | |  ||||      |  |  ||| | || | |   | | || |||  ||
    GGAT---CACGATCTGGGCCATGAGCATCCCGAGCACGAACTCCAGCAGCCGCACCGGCGGGAACATGTAGA
            540        550        560        570        580        590        600

         200        210        220        230        240        250        260
    TGATAGATGTCTGTCTCA-CTGCGTTGTTTAAGGCTCCGCAGGATGAGGCCTTGGCCAGGTCAGTGCTCCAC
    |||   |   |  | || |||||   | |  || |||   ||  | |||   || | || | |     ||
    TGA---ACCACCACCGCAGCTGCGGCATGTCCGGGTCCCACGGCAGGGGC---GGGCTGGCCGG-----CAG
              610        620        630        640        650            660

        270        280        290        300        310        320        330
    AAGCAGCAGGTCCTGGGGACTGGGGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCGCCGCGTCCT
     |||||   | || ||||  ||| ||||| || |||  |  | ||  ||||    | || |||||    |
    CAGCAGTGCGACCAGGGGGATGGAGACGGCGGCCAC--GGACACCGCGGCG----GCCCACCGCCAGAGC--
          670        680        690          700        710            720

       340        350        360        370        380        390        400
    CTGCCAGTGCAGGCACCAG-GCAG-CGGCCAGCAGCAGGAGGCCCACGG-GCAGCAG--CAG-TAGGAGGAG
    | | | || | | |    | | ||  ||| |  || |   ||  || || | || ||  ||| | | |||||
    CGGTCCGTACGGACCTTGGTGAAGAAGGCGAAGAGGAACGGGAACATGGCGTAGAAGAACAGCTCGCAGGAG
      730        740        750        760        770        780        790

           410        420        430        440        450        460
    -CAGAGGGGGCTGCGGGGCTGTCGGGGCTGTGGCCTCCAGG-GGCCGGGGACTCCATG--GGGGT-GGCAG-
       ||    ||  | ||| |   |  || ||   |||  || || ||||||  |||||   || |  ||||
    AGCGACCACGCCACCGGGTTCATGCTGCCGT--ACTCGTGGTGGTCGGGGA-ACCATGCCTGGATCAGCAGC
    800        810        820        830        840        850        860
```

```
  470          480        490        500        510        520        530
-GGT---TGAGGAGTCGGGCTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATCCAGGGCT
 |||   |||| |||| | |  |||| | | |    ||| |  |  |  |  | | | |  | |  |  |
AGGTTCGTGAGCAGTCCGTCCCACACCG-ATC----GGCCCATGTTGGGCTCGTTGAGGGCCAGCACGATCA
 870        880        890             900        910        920        930

  540        550          560        570        580        590        600
TCAGCGCCACCAGCTGCTCGG--AGGTCTCCTGC-AGGAGGCGGGAGATGTTGGTCTGGACGAAGCGAAG-A
 ||||| ||||||| || |||  |||  | | || || | ||| | ||   ||  |  | ||    ||||
GCAGCGTCACCAGCAGCACGGGCAGG--TGCAGCGAGTACGCGCG-GACCGTGCGC-CGCCG-CCAGAAGTT
    940        950        960        970        980        990       1000

    610           620        630        640        650        660
CAGCTGGG----GGGGGGCTGAAAGGCACATTTGGTGACAAAGTGTATCTCCGTGTTCACG--CG----CTC
|| || ||    |  |||| ||   || ||  ||  ||| ||      |   | ||   | |  ||    |
CACCTTGGACTTGTCGGGCAGACCCGCCCAGGTGAGGACGAAACCGCTGAGCATGAAGAAGAACGAGACCGT
      1010       1020       1030       1040       1050       1060       1070

  670        680        690        700        710          720          730
CAGCAAGCCTTGCATCTTGGACCCAGCGACAGTCTTGAGCCGCTCCATCCA--GCGCTGTGCCAG--GACCA
||||  |||  ||   ||      |||||  |  ||  || |||   ||  |   || ||| | |  || |
CAGCGGGCCCAGCCGGTT-----CAGCGGGA-ACTGCAG-CGCGGAATTGATCTCGGTGTTCTTGAAGAACG
    1080       1090            1100       1110       1120       1130

       740        750        760        770          780        790
GCCG---CCAGAGGCCCCCGCAGAGCTCCTCGTCCTGCAG-GTTG-GAGGC--CACGGTGA-CTGGGTA--A
|| |   | | |        |  ||  | |  ||   || | | ||  | ||  | ||  || |  || |  |
GCTGTGTCGATATATGGGAGGTGA-ATACCAGTAGAGCGGAGATGAAACGCATCCCGCCGAGCGCGGGAAGA
1140       1150       1160       1170       1180       1190       1200

  800        810          820        830        840        850        860
TCTTGAAGCAGGTAGTCA---GACAGCTCACGGATTTTGACAGCGAAGTCGGAGGAGATGGGGCTGTGTTGG
| ||   |||||  | ||   |  | ||| |||      ||| |    || |||| ||||  | ||  |  |
TGTTTCGGCAGG--GGCATGGGTGACCTCGCGG---CGGACGGGTGGGTGGGAGCAGAT--GTCTAAG--GC
1210       1220       1230       1240       1250       1260       1270

    870        880        890        900        910        920        930
AAGGAGCAGTC-CTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGGAG-GAGATAGGTTGTTGGGC
|  | || | | | |||   ||  | ||  |   |  || ||    ||| |  | | || ||  || || |
ACCGCGCGGCCGCCGGG---CAAGGCGTATCTGACAAAG-AGTGTGAGCCGCCGCGCGACAG--TGGTGTGT
     1280       1290       1300       1310       1320       1330

     940        950        960        970        980        X
TCC--AGGCTG-GCG-CCAGC-ACTGTCATTTCG--GCCGGGGGCCCCCGGAATT--CCGGCCG
|||    ||||  || || || || ||  |||||   |||||  | | |||| ||   |  |||
TCCCACCGCTGCACGGCCGGCGACGGTGGTTTCGTCACCGGGCGACGCGGATTTAATCATCCGGAAACACGG
1340       1350       1360       1370       1380       1390       1400

   A
1410
```

5. US-08-162-407-5' (1-988)
   Q29269 Human calcium channel 27980/11.

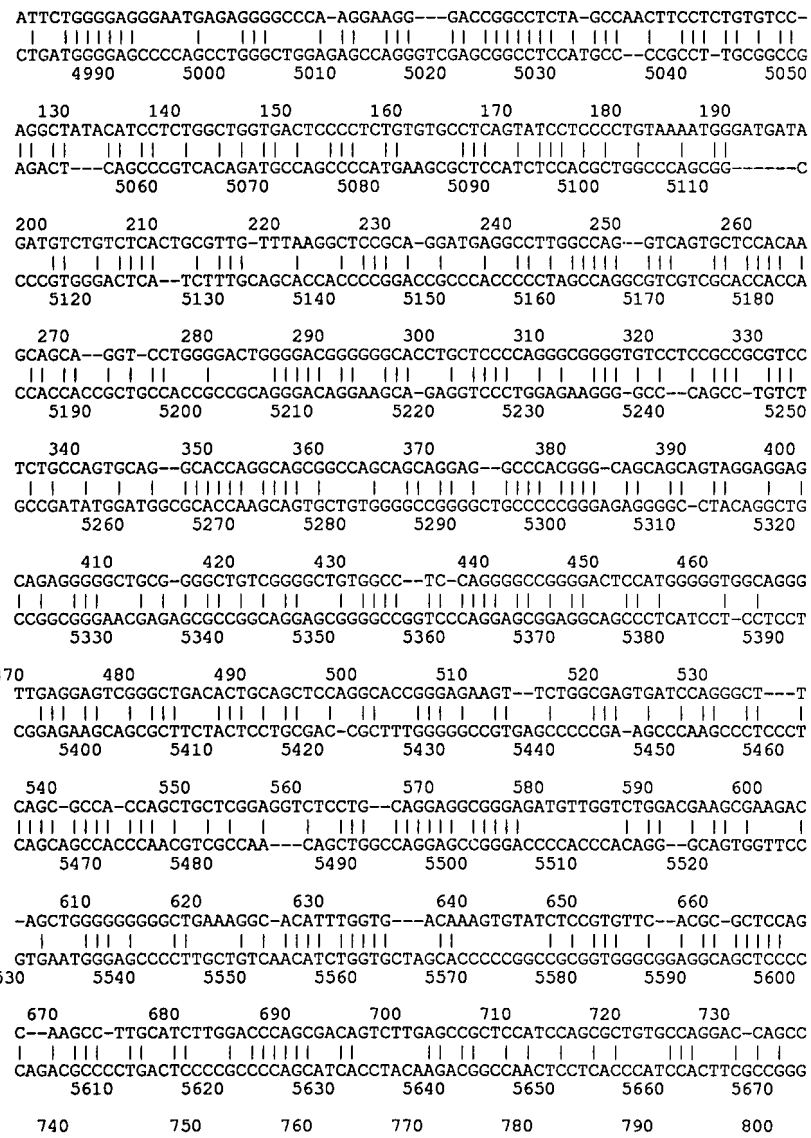
```
ID   Q29269 standard; DNA; 6232 BP.
AC   Q29269;
DT   03-MAR-1993  (first entry)
DE   Human calcium channel 27980/11.
KW   Plasmid pR14-5.3.3.1; Ca-flux assay; ss.
```

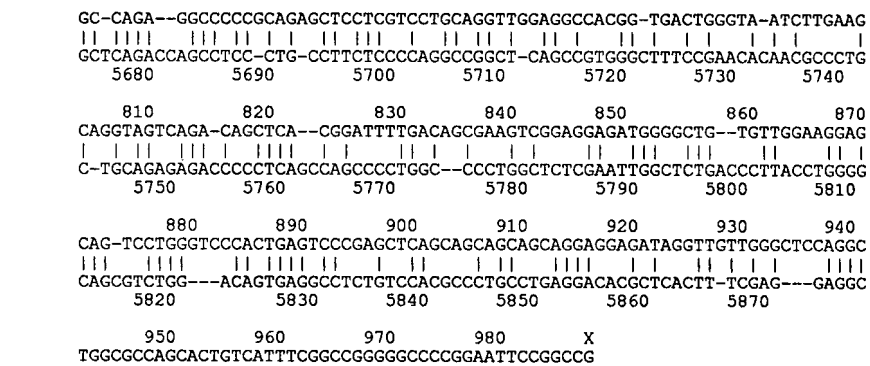


```
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   polyA_signal    6215..6220
FT   /*tag= a
FT   repeat_unit     1..252
FT   /*tag= b
FT   /standard_name= Alu_repeat
FT   /note= "possible cloning artefact"
FT   CDS             253..6048
FT   /*tag= c
FT   /note= "amino acids 358 to C-terminus
FT   i.e. Domains II to IV"
FT   misc_difference 3746
FT   /*tag= d
FT   /note= "undefined"
PN   EP-507170-A.
PD   07-OCT-1992.
PF   23-MAR-1992; 104970.
PR   04-APR-1991; DE-110785.
PA   (FARB ) BAYER AG.
PI   Franz J, Rae P, Unterbeck A, Weingaertner B;
DR   WPI; 92-333446/41.
DR   P-PSDB; R27649.
PT   Cloned human neuronal calcium channel sub-types - useful in
PT   calcium flux assays to screen for neurone-specific calcium
PT   channel ligands
PS   Claim 2; Page 63-77; 101pp; German.
CC   Human neuroblastoma cell line, hippocampus, frontal and temporal
CC   cortex and visual cortex cDNA banks were screened with a probe
CC   containing carp skeletal muscle Ca-channel cDNA. The cDNA clone
CC   pR14-5.3.3.1 overlaps with clone p1247-14.1.1.1 (see Q29263). The
CC   following differences are observed between the two sequences
CC   (nucleotide and position in pR14-5.3.3.1 given in brackets):
CC   1.Cytosine at position 520 (T: 3507); no change in deduced amino
CC   acid sequence. 2.Cytosine at position 775 (G; 3768); no change in
CC   deduced AA sequence. 3.Cytosine at position 1617 (T;4611).
CC   4.Adenosine at position 2360 (G; 5353). 5.deletion of 6 nucleotides
CC   at position 708 (CGGAAA; 3695-3700). 6. deletion of an Adenosine
CC   residue at position 1013 which leads to a stop codon at position
CC   1028-1030. 7. at position 3240 there are a further 2199 nucleotides
CC   of the 3'UTR which are absent from pR14-5.3.3.1. (The deletion of
CC   Adenosine at position 1013 is thought to be a cloning artefact).
CC   The sequence can be inserted into a eukaryotic expression vector for
CC   use in transforming suitable host cells. Cell lines producing human
CC   neuronal calcium channel proteins can be used for screening for Ca
CC   channel ligands (agonists or antagonists). See also Q29259-Q29275.
SQ   Sequence   6232 BP;    1250 A;    1914 C;    1827 G;    1240 T;
SQ   1 Others;
```

```
Initial Score    =    144  Optimized Score  =    442  Significance =  8.51
Residue Identity =    49%  Matches          =    518  Mismatches   =   423
Gaps             =    104  Conservative Substitutions               =     0
```

```
             X       10        20         30        40         50
             CGGGGGCTTTGTTCAGAGTTGG--GTAAATGGGGCCGGTGCTG-GATTCTGGCGTTGCCTCC
               |||||  |    ||   |||  ||  |  |  | | ||| | |  |    ||  |   ||
   AGAGATCCCTGTGGGGCGGTCAGGAGCACTGGCTGTGGA-CGTTCAGATGCAGAGCATAACCCGGAGGGGCC
        4920      4930      4940       4950      4960       4970      4980

  60        70        80         90          100        110        120
```





6. US-08-162-407-5' (1-988)
Q37817 Sequence encoding the alpha 1B-1 human calcium cha

ID Q37817 standard; cDNA; 7362 BP.
AC Q37817;
DT 30-JUN-1993 (first entry)
DE Sequence encoding the alpha 1B-1 human calcium channel subunit.
KW Human calcium channel subunit; diagnosis; agonist; antagonist;
KW Lambert Eaton syndrome; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 144..7163
FT /*tag= a
PN WO9304083-A.
PD 04-MAR-1993.
PF 14-AUG-1992; U06903.
PR 15-AUG-1991; US-745206.
PR 10-APR-1992; US-868354.
PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,
PI Williams ME;
DR WPI; 93-093936/11.
DR P-PSDB; R33549.
PT DNA encoding specific human calcium channel sub-units - used for
PT identifying calcium channel agonists and antagonists and
PT diagnosing Lambert Eaton syndrome
PS Disclosure; Page 111-120; 150pp; English.
CC DNA encoding the alpha 1B subunit was isolated by screening a
CC human basal ganglia cDNA library with fragments of the rabbit
CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.
CC A portion of one of the positive clones was used to screen an IMR32
CC cell cDNA library. Clones that hybridized to the basal ganglia
CC DNA prove were used to further screen an IMR32 cell cDNA library
CC to identify overlapping clones that in turn were used to screen a
CC human hippocampus cDNA library. In this way, a sufficient series of
CC clones to span nearly the entire length of the nucleotide sequence

```
CC   encoding the human alpha 1B subunit was obtained. PCR amplification
CC   of specific regions of the IMR32 cell alpha 1B mRNA yielded
CC   additonal segments of the alpha 1B coding sequence. A full-length
CC   alpha 1B DNA clone was constructed by ligating portions of the
CC   partial cDNA clones (see Q37817, Q37818). Alpha 1B-1 and alpha
CC   1B-2 are derived by alternative splicing of the alpha 1B subunit
CC   transcript.
SQ   Sequence   7362 BP;    1446 A;    2276 C;    2214 G;    1426 T;

Initial Score    =    143  Optimized Score  =    442  Significance =  8.44
Residue Identity =    49%  Matches          =    518  Mismatches   =   423
Gaps             =    104  Conservative Substitutions              =     0

             X       10        20          30        40         50
             CGGGGGCTTTGTTCAGAGTTGG--GTAAATGGGGCCGGTGCTG-GATTCTGGCGTTGCCTCC
               |||||  |    ||   |||  ||  |  |  | | ||| | |  |    ||  |   ||
   AGAGATCCCTGTGGGGCGGTCAGGAGCACTGGCTGTGGA-CGTTCAGATGCAGAGCATAACCCGGAGGGGCC
    6030      6040      6050      6060       6070      6080      6090

  60        70        80         90         100       110       120
   ATTCTGGGGAGGGAATGAGAGGGGCCCA-AGGAAGG---GACCGGCCTCTA-GCCAACTTCCTCTGTGTCC-
     |  |||||||     |   |||    | ||  |||   || |||||||  | |||  |  |||  || | ||
   CTGATGGGGAGCCCCAGCCTGGGCTGGAGAGCCAGGGTCGAGCGGCCTCCATGCC--CCGCCT-TGCGGCCG
   6100      6110      6120      6130      6140      6150       6160

     130       140       150       160       170       180       190
   AGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGTATCCTCCCCTGTAAAATGGGATGATA
   || ||   || ||  |  | | || |  |||  ||     |||  | ||| | |  |    |  ||
   AGACT---CAGCCCGTCACAGATGCCAGCCCCATGAAGCGCTCCATCTCCACGCTGGCCCAGCGG------C
    6170          6180      6190      6200      6210      6220

   200       210       220       230       240       250       260
   GATGTCTGTCTCACTGCGTTG-TTTAAGGCTCCGCA-GGATGAGGCCTTGGCCAG--GTCAGTGCTCCACAA
      ||  | ||||   | |||     |   | ||| |  |    |  || | ||||| |||   || ||||
   CCCGTGGGACTCA--TCTTTGCAGCACCACCCCGGACCGCCCACCCCCTAGCCAGGCGTCGTCGCACCACCA
  6230      6240      6250      6260      6270      6280      6290

     270       280       290       300       310       320       330
   GCAGCA--GGT-CCTGGGGACTGGGGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTCCGCCGCGTCC
    || ||  | | ||    |      ||||| ||  |||   | ||||  | |  ||| | |  | |||  |||
   CCACCACCGCTGCCACCGCCGCAGGGACAGGAAGCA-GAGGTCCCTGGAGAAGGG-GCC--CAGCC-TGTCT
  6300      6310      6320      6330      6340      6350      6360

       340       350       360       370       380       390       400
   TCTGCCAGTGCAG--GCACCAGGCAGCGGCCAGCAGCAGGAG--GCCCACGGG-CAGCAGCAGTAGGAGGAG
    | |  | |  |  |||||| |||| |     |  || || |  |||| ||||  ||  ||  ||    | |
   GCCGATATGGATGGCGCACCAAGCAGTGCTGTGGGGCCGGGGCTGCCCCCGGGAGAGGGGC-CTACAGGCTG
    6370      6380      6390      6400      6410      6420      6430

         410       420       430       440       450       460
   CAGAGGGGGCTGCG-GGGCTGTCGGGGCTGTGGCC--TC-CAGGGGCCGGGGACTCCATGGGGGTGGCAGGG
   | |  |||   | | | || | | ||   | ||||  || |||| || | ||   || | |   | |
   CCGGCGGGAACGAGAGCGCCGGCAGGAGCGGGGCCGGTCCCAGGAGCGGAGGCAGCCCTCATCCT-CCTCCT
   6440      6450      6460      6470      6480      6490      6500

  470       480       490       500       510       520       530
   TTGAGGAGTCGGGCTGACACTGCAGCTCCAGGCACCGGGAGAAGT--TCTGGCGAGTGATCCAGGGCT---T
     ||| ||  | |||   ||| | ||  |  ||  ||| |  ||   |   |||  |  | ||  ||   |
   CGGAGAAGCAGCGCTTCTACTCCTGCGAC-CGCTTTGGGGGCCGTGAGCCCCCGA-AGCCCAAGCCCTCCCT
  6510      6520      6530      6540      6550      6560      6570
```



```
      540       550       560       570       580       590       600
   CAGC-GCCA-CCAGCTGCTCGGAGGTCTCCTG--CAGGAGGCGGGAGATGTTGGTCTGGACGAAGCGAAGAC
   |||| |||| ||| |  | | |  |   |  |||  |||||| ||||          | ||  | || |    |
   CAGCAGCCACCCAACGTCGCCAA---CAGCTGGCCAGGAGCCGGGACCCCACCCACAGG--GCAGTGGTTCC
  6580      6590      6600      6610      6620      6630      6640

      610       620       630       640       650       660
   -AGCTGGGGGGGGGCTGAAAGGC-ACATTTGGTG---ACAAAGTGTATCTCCGTGTTC--ACGC-GCTCCAG
     |  ||| |   ||    | | |||| |||||    ||         | | |||  |  | || |||||
   GTGAATGGGAGCCCCTTGCTGTCAACATCTGGTGCTAGCACCCCCGGCCGCGGTGGGCGGAGGCAGCTCCCC
     6650      6660      6670      6680      6690      6700      6710

   670       680       690       700       710       720       730
   C--AAGCC-TTGCATCTTGGACCCAGCGACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGAC-CAGCC
   |  | |||  ||  ||   | ||||||  ||       || ||  | |  |  | |  |||   | | |
   CAGACGCCCCTGACTCCCCGCCCCAGCATCACCTACAAGACGGCCAACTCCTCACCCATCCACTTCGCCGGG
   6720      6730      6740      6750      6760      6770      6780

     740       750       760       770       780       790       800
   GC-CAGA--GGCCCCCGCAGAGCTCCTCGTCCTGCAGGTTGGAGGCCACGG-TGACTGGGTA-ATCTTGAAG
   || |||   ||| || | |  || |||  |  || || |  || |   || |  | |  | | | |   |
   GCTCAGACCAGCCTCC-CTG-CCTTCTCCCCAGGCCGGCT-CAGCCGTGGGCTTTCCGAACACAACGCCCTG
 6790      6800      6810      6820      6830      6840      6850

      810       820       830       840       850       860       870
   CAGGTAGTCAGA-CAGCTCA--CGGATTTTGACAGCGAAGTCGGAGGAGATGGGGCTG--TGTTGGAAGGAG
   |  | ||  ||| |  |||| | |    || | |  | | |    || ||| ||| |  ||   || | |
   C-TGCAGAGAGACCCCCTCAGCCAGCCCCTGGC--CCCTGGCTCTCGAATTGGCTCTGACCCTTACCTGGGG
   6860      6870      6880      6890      6900      6910      6920

          880       890       900       910       920       930       940
   CAG-TCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGATAGGTTGTTGGGCTCCAGGC
   |||  ||||    || |||| ||   |  ||    | ||  ||||   | |  || | | |    ||||
   CAGCGTCTGG---ACAGTGAGGCCTCTGTCCACGCCCTGCCTGAGGACACGCTCACTT-TCGAG---GAGGC
   6930          6940      6950      6960      6970      6980          6990

        950       960       970       980       X
   TGGCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGGAATTCCGGCCG
   ||  |||| ||      | |  |||||      || ||| |||| |
   TGTGGCCACCA-----ACTCGGGCCGCTCCTCCAGGACTTCCTACGTGTCCTCCCTG
        7000          7010      7020      7030  X   7040
```

7. US-08-162-407-5' (1-988)
   Q54631 Human oncogene bcl-2 coding sequence.

```
ID   Q54631 standard; cDNA to mRNA; 5086 BP.
AC   Q54631;
DT   23-JUN-1994  (first entry)
DE   Human oncogene bcl-2 coding sequence.
KW   Cell death; senescence; programmed cell death; ced-9; myocardial
KW   infarction; stroke; brain injury; neurodegenerative disease;
KW   muscular degenerative disease; ageing; hypoxia; ischaemia; toxaemia;
KW   infection; hair loss; neoplasia; cancer; ced-3; ced-4; bcl-2;
KW   oncogene; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1459..2178
FT   /*tag=  a
```

```
FT   /product= Bcl-2.
PN   WO9325683-A.
PD   23-DEC-1993.
PF   14-JUN-1993; U05651.
PR   12-JUN-1992; US-898933.
PR   10-AUG-1992; US-927681.
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI   Hengartner M,  Horvitz HR;
DR   WPI; 94-007540/01.
DR   P-PSDB; R47344.
PT   Caenorhaboitis elegans cell death-protective gene - used to
PT   develop agents for preventing cell death or for reducing
PT   population of cells
PS   Disclosure; Page 61-64; 112pp; English.
CC   The protein product of the human oncogene bcl-2 was found to have a
CC   similar sequence to the ced-9 protein.  ced-9 is essential for
CC   C. elegans development and apparently functions by protecting cells
CC   during development from programmed cell death.  ced-9 was shown to
CC   function by antagonising the activities of cell death genes ced-3
CC   and ced-4.  The ced-9 gene can be used for developing agents for
CC   treating a condtion characterised by increased cell death such as
CC   myocardial infarction, stroke, traumatic brain injury,
CC   neurodegenerative disease, muscular degenerative disease, ageing,
CC   hypoxia, ischaemia, toxaemia, infection or hair loss.  It can also
CC   be used for reducing a population of cells in the treatment of
CC   neoplastic growth cancerous tissue, infected cells or autoreactive
CC   immune cells.
SQ   Sequence   5086 BP;    1262 A;    1222 C;    1288 G;    1314 T;

Initial Score     =    141  Optimized Score  =    321  Significance =  8.31
Residue Identity  =    47%  Matches          =    381  Mismatches   =   327
Gaps              =     88  Conservative Substitutions               =     0

  230       240       250       260       270       280       290       300
    GCAGGATGAGGCCTTGGCCAGGTCAGTGCTCCACAAGCAGCAGGTCCTGGGGACTG-GGGACGGGGGGCACC
              ||    ||| |  |    |||| |  || ||  | ||     | | |  |  |    | |  |
              GC----GCCCGCCC----CTCCGC--GCCGCCTGCCCGCCCGCCCGCGCGCTCCCGCCCGC
              X           10              20        30        40        50

          310       320       330       340       350       360       370
    TGCTCCCCAGGGCGGGGTGTC-CT-CCGCCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCA
     |||| ||  |||   |  || || |||||||  || |||||||| | |||  || ||  ||  || |   |
    CGCTCTCCGTGGCCCCGCCTCGCTGCCGCCGCCGCCGCTGCCAGCGAAGGTGCCGGG--GC-TCCGGGCCCT
          60        70        80        90       100       110       120

          380       390       400       410       420       430       440
    GGAGGCCCACGGGCAGCAGCAGTAGGAGGAGCAGAGGGGGCTGCGGGGCTGTCGGGGCTGTGGCCTCCAGGG
        |||  ||| |  ||||  |    |||||  |  | || |  |||  ||| ||   | |||  ||   |
    CCCTGCCGGCGGCCGTCAGCGCT---CGGAGCGAACTGCGC-GACGGGAGGTCCGG---GAGGCGACCGTAG
          130       140         150       160       170         180

         450       460       470       480       490       500       510
    GCCGGGGACTCCATGGGGGTGGCAGGGTTGAGGAGTCGGGCTGACACTGCAGCTCCAGGCACCGGGAGAAGT
     |  | | | ||  |  ||   || ||  ||||||   || || | | ||  |   ||||   | | |  |
    TC--GCGCCGCCGCGCAGG-ACCA-GGAGGAGGAGAAAGGGTG-CGCAGC--CCGGAGGCGGGGTGCGCCGG
       190       200         210       220       230         240       250

      520       530       540       550       560       570
    TCTGGCG-AG-TGATCCAGGGCTTCAGCGCCACCAGCTGCTCGGAG--GTC-TCC--TGCAGG--AGGCGGG
    |  || | ||  ||    ||| | ||| |     | ||  ||| ||  ||| |||  |  |||  | | |||
```



```
    TGGGGTGCAGCGGAAGAGGGGGTCCAGGGGGGAGAACT--TCGTAGCAGTCATCCTTTTTAGGAAAAGAGGG
          260       270       280       290       300       310       320

       580       590       600       610       620       630       640
    ---AGATGTTGGTCTGGACGA--AGCGAAGACAGCTGGGGGGGGGGCTGAAAGGCACATTTGGTGACAAAGTG
       | ||        ||   | |  | |     |  |          || | |   ||||    | |      |
    AAAAAATAAAACCCTCCCCCACCACCTCCTTCTCCCCACCCCTCGCCGCA--CCACACACAGCGCGGGCTTC
          330       340       350       360       370       380       390

       650       660       670       680       690       700
    TATCTCCGTGTTCACGCGCTCCA-GCAAG-CCTTGCATCTTGGACCCAGCGACAGTCTTGA----GCCGCT-
    || | |   |  |  |||  ||| ||  | |||  | || |  | |||||  |  |   ||    ||   |
    TAGCGCTCGGCACCGGCGGGCCAGGCGCGTCCTGCCTTCATTTATCCAGCAGCTTTTCGGAAAATGCATTTG
          400       410       420       430       440       450       460

710       720       730       740       750       760       770
    CCATCCAGCGCTGTGCCAG--GACCA--GCCGCCAGAGGCCCCCGCAGAGCTCCTCGTCCTGCAGGTTGGAG
    |  | | | | |    |||  ||| |   | |||   | ||||  ||      || ||||||  ||  |
    CTGTTCGGAGTTTAATCAGAAGACGATTCCTGCCTCCGTCCCCGGC-TCCTTCATCGTCC--CATCTCCCCT
        470       480       490       500       510       520       530

    780       790       800       810       820       830       840
    GCCACGGTGACTGGGTAATCTTGAAGCAG----GTAGTCAGACAGCTCACGGATTTTGACAGCG-AAGT---
    | | |  |  ||||| |  | |||||| |     || |  ||| |  |||  |  | || | |||   ||
    GTCTCTCT-CCTGGGGAGGCGTGAAGCGGTCCCGTGGATAGAGA-TTCA-TGCCTGTGTCCGCGCGTGTGTG
         540       550       560       570       580       590       600

           850       860       870       880       890       900       910
    CG-GAGGAGATGGGGCTGTGTTGGAAGGAGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGCAGCAGCA-G
    || | | | |    || | |  ||    | ||   | ||     |   || | |||  || | |     |
    CGCGCGTATAAATTGCCGAGAAGGGGAAAACATCACAGGACTTCTGCGAATACCGGACTGAAAATTGTAATT
          610       620       630       640       650       660       670

          920       930       940       950       960       970
    CAGGAGGAGATAGGTTGTTGGGCTCCAGGCTGGCGCCAGCAC-TGTCAT-TTCGGCCGGGGGCCC--CGGA-
    ||   |  |    |  | |  ||   |       | | ||| | ||  || | |||  |||   ||  ||||
    CATCTGCCG--CCGCCGCT--GCCAAAAAAAAAACTCGAGCTCTTGAGATCTCCGGTTGGGATTCCTGCGGAT
       680       690       700       710       720       730       740

       980       X
    ----ATTCCGGCCG
        |||  | |
    TGACATTTCTGTGAAGCAGAAGTC
          750    X  760
```

8. US-08-162-407-5' (1-988)
   N81292 Sequence of bcl-2 cDNA corresp. to the 5.5 kb tran

```
ID   N81292 standard; cDNA; 5105 BP.
AC   N81292;
DT   17-DEC-1990  (first entry)
DE   Sequence of bcl-2 cDNA corresp. to the 5.5 kb transcript encoding
DE   bcl-2-alpha
KW   B-cell neoplasm; diagnosis; follicular lymphomas; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1459..2178
FT   /*tag=  a
PN   EP-252685-A.
```

```
PD   13-JAN-1988.
PF   02-JUL-1987; 305863.
PR   09-JUL-1986; US-883687.
PA   (WIST-) Wistar Corp.
PI   Tsujimoto Y, Croce CM;
DR   WPI; 88-008633/02.
DR   P-PSDB; P80987.
PT   Detection of B-cell neoplasms -
PT   by extn. of proteins or RNA from B-cells and quantitation using
PT   specific antibody or DNA probe
PS   Claim 8; Fig 2A-2D; 23pp; English.
CC   A human bcl-2 gene substantially free of introns is claimed. Also claimed
CC   is a substantially pure preparation of a protein having an N-terminal end
CC   encoded by the first exon of the human bcl-2 gene wherein said protein is
CC   bcl-2-alpha having about 239 (P80987) or 205 (P80988) AA residues. B-cell
CC   neoplasms which are associated with t(14;18) chromosome translocations
CC   cause an increase in the expression of both mRNA and the protein prods.
CC   of the bcl-2 gene. This is used to detect B-cell neoplasms including
CC   follicular lymphomas as well as other lymphomas. Bacterial isolates
CC   available as ATCC 67147 and 67148 can be used to express gene prods.
CC   alpha (n81292) and beta (n91293) resp. in bacteria.
SQ   Sequence   5105 BP;   1281 A;   1225 C;   1286 G;   1313 T;

Initial Score     =    140  Optimized Score  =    320  Significance =  8.24
Residue Identity  =    48%  Matches         =    386  Mismatches   =   312
Gaps              =     95  Conservative Substitutions              =     0

     230       240       250       260       270       280       290
    TCCGCAGGATGAGGCCTTGGCCAGGTCAGTGCTCCACAAGCAGCAGGTCCTGGGGACT-GGGGACGGGGGGC
              | | ||   |||   || | || | |   | || | ||   |  ||   |  ||  |  |
              GCGCCC---GCC-CCTCCGCGC-CGCCTGCCCGCCCGCCCGCGCTCCCGCCCGCCGCTC
              X           10        20        30        40        50

    300       310       320       330       340       350       360
    ACCTGCTCCCCAGGGCGGGGTGTCCTCCGCCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGC
     ||    ||||  | || | || || ||||||   || ||||||| | |||  || ||  ||  || |   |
    TCCGTGGCCCC--GCCGCGCTG-CCGCCGCCG---CCGCTGCCAGCGAAGGTGCCGGG--GC-TCCGGGCCC
     60          70        80        90        100       110       120

  370       380       390       400       410       420       430       440
    AGGAGGCCCACGGGCAGCAGCAGTAGGAGGAGCAGAGGGGGCTGCGGGGCTGTCGGGGCTGTGGCCTCCAGG
         |||  ||| |  |||| |    |||||  |  | || |  |||  ||| ||   | |||  ||
    TCCCTGCCGGCGGCCGTCAGCGCT---CGGAGCGAACTGCGC-GACGGGAGGTCCGG---GAGGCGACCGTA
          130       140       150       160       170       180

        450       460       470       480       490       500       510
    GGCCGGGGACTCCATGGGGGTGGCAGGGTTGAGGAGTCGGGCTGACACTGCAGCTCCAGGCACCGGGAGAAG
    | |  | | | ||  |  ||   || ||  ||||||   || || | | ||  |   ||||   | | |  |
    GTC--GCGCCGCCGCGCAGG-ACCA-GGAGGAGGAGAAAGGGTG-CGCAGC--CCGGAGGCGGGGTGCGCCG
        190       200       210       220       230       240       250

        520       530       540       550       560       570
    TTCTGGCG-AG-TGATCCAGGGCTTCAGCGCCACCAGCTGCTCGGAG--GTC-TCC--TGCAGG--AGGCGG
     |  || | ||  ||    ||| | ||| |      | ||  ||| ||  ||| ||| |  |||  | | ||
    GTGGGGTGCAGCGGAAGAGGGGGTCCAGGGGGGAGAACT--TCGTAGCAGTCATCCTTTTTAGGAAAAGAGG
           260       270       280       290       300       310       320

        580       590       600       610       620       630       640
    G---AGATGTTGGTCTGGACGA--AGCGAAGACAGCTGGGGGGGGGCTGAAAGGCACATTTGGTGACAAAGT
    |   | ||      ||   | |  | |     |  |          || | |   ||||    | |      |
```



```
    GAAAAAATAAAACCCTCCCCCACCACCTCCTTCTCCCCACCCCTCGCCGCA--CCACACACAGCGCGGGCTT
          330       340       350       360       370       380       390

       650       660       670       680       690       700
    GTATCTCCGTGTTCACGCGCTCCA-GCAAG-CCTTGCATCTTGGACCCAGCGACAGTCTTGA----GCCGCT
     || | |   |  |  |||  ||| ||  | |||  | || |  | |||||  |  |   ||    ||   |
    CTAGCGCTCGGCACCGGCGGGCCAGGCGCGTCCTGCCTTCATTTATCCAGCAGCTTTTCGGAAAATGCATTT
          400       410       420       430       440       450       460

  710       720       730       740       750       760       770
    -CCATCCAGCGCTGTGCCAG--GACCA--GCCGCCAGAGGCCCCCGCAGAGCTCCTCGTCCTGCAGGTTGGA
     |  | | | | |   |||  ||| |   | |||   | |||| ||     || ||||||  ||  |
    GCTGTTCGGAGTTTAATCAGAAGACGATTCCTGCCTCCGTCCCCGGC-TCCTTCATCGTCC--CATCTCCCC
        470       480       490       500       510       520       530

   780       790       800       810       820       830       840
    GGCCACGGTGACTGGGTAATCTTGAAGCAG----GTAGTCAGACAGCTCACGGATTTTGACAGCGAAGTCGG
     | | |  |||||  |  | ||||||  |    || |  ||| |  |||  |  | || | |||    ||
    TGTCTCTCT-CCTGGGGAGGCGTGAAGCGGTCCCGTGGATAGAGA-TTCA-TGCCTGTGTCCGCG----CG-
          540       550       560       570       580       590

      850       860       870       880       890       900       910
    AGGAGATGGGGCTGTGTTGGAAGGAGCAGTCCTGGGTCCCAC-TGAGTCCCGAGCTCAGCAGCAGCAGCAGG
      | |   | || || |   ||   || |    |||    || | |  |  ||  || || | |  | | |
    -TGTGTGCGCGC-GTAT---AAATTGCCGAGAAGGGGAAAACATCA--CAGGACTTCTGC-GAA-TACCGGA
      600         610         620         630         640         650

     920       930       940       950       960       970
    AGGAGATAGGTTGTTGGGCTCCAG--GCTGGCGCCAGCACTGTCATTTCG-GC-CGGGGG----CCC--CGG
      || |   ||  ||   || | |  || |  ||||   |      |  ||| || |  | |    |||   ||
    CTGAAAATTGTAATTCATCTGCCGCCGCCGCTGCCA--AAAAAAAACTCGAGCTCTTGAGATCTCCCGTTGG
   660       670       680       690       700       710       720

    980       X
    AATTCCGGCCG
     ||||| || |
    GATTCCTGCGGATTGACATTT
   730       740
```

9. US-08-162-407-5' (1-988)
   Q05884 Plasmid clone 409.2.

```
ID   Q05884 standard; DNA; 633 BP.
AC   Q05884;
DT   11-JAN-1991  (first entry)
DE   Plasmid clone 409.2.
KW   Basic fibroblast growth factor; tissue regeneration; infarction; ss.
FH   Key             Location/Qualifiers
FT   CDS             1..633
FT   /*tag=  a
FT   /label=basic fibroblast growth factor
PN   FR2642086-A.
PD   27-JUL-1990.
PF   26-JAN-1989; 000973.
PR   26-JAN-1989; FR-000973.
PA   (SNFI ) SANOFI SA.
PI   Caput D, Ferrara P, Kaghad M;
DR   WPI; 90-277408/37.
DR   P-PSDB; R07076.
```

```
PT   New recombinant gene encoding basic fibroblast growth factor - in
PT   new high mol. wt. form, useful e.g. for stimulating tissue
PT   regenerating or treating infarction
PS   Disclosure; fig 8; 43pp; French.
CC   This recombinant gene encodes an extended basic fibroblast growth
CC   factor (bFGF) of 210 amino acids.  It stimulates growth of meso-
CC   dermal and neuroectodermal cells.  It is thus potentially useful
CC   e.g. for regenerating damaged tissues, and for treating myocardial
CC   infarctions, Parkinsons disease and Alzheimers disease.  It can be
CC   produced on a large scale using recombinant DNA methods without
CC   risk of contamination.  See also Q05883.
SQ   Sequence   633 BP;    134 A;    172 C;    211 G;    116 T;

Initial Score    =    137  Optimized Score  =    238  Significance =  8.03
Residue Identity =    50%  Matches          =    286  Mismatches   =   206
Gaps             =     69  Conservative Substitutions               =     0

      460     X 470         480         490         500         510         520
    TGGGGGTGGCAGGGTTGA--GGAGTCGGGCTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTG
              | || ||| |  |||| | |  | | || |  || ||   ||||||  | |  || |   |
              ATGGGTGACCGTGGTCGTGGTCGCGC-GCTG--CCGGG---CGGGAG--GCTGGGGGGCCGG
              X       10          20          30              40          50

      530         540         550         560         570         580         590
    ATCCAGGGCTTCAGCGCCAC-CAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGTTGG-TCTGGACGA
      || |||   | | ||| |  ||| | |||||| |    || || | |||| || |  || ||    ||
    GGCCGGGG---CCGTGCCCCGGAGCGGGTCGGAGG-CCGGGGCCGGGGCCGGGGGACGGCGGCTCCCCGCGC
        60          70          80          90         100         110         120

      600         610         620         630         640         650         660
    AGCGAAGACAGCTGGGGGGGGGCTGAAAGGCACATTTGGTGACAAAGTGTATCTCCGTG-TTCA-CGCGCTC
     ||     | ||| ||||    | |   ||| |    || ||| | | | | |   | |  ||| | ||||
    GGCTCCAGCGGCTCGGGGATCCCGGCCGGGCCCCGCAGG-GACCATG-GCAGCCGGGAGCATCACCACGCT-
         130         140         150         160         170         180         190

      670         680         690         700         710         720         730
    CAGCAAGCCTTGCATCTTGGACCCAGCGACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCCAGGACC---AG
      ||  |||||||  |  |||    ||| |||    | |||  |||  || |  |  | | ||||||   ||
    --GCCCGCCTTGC-CCGAGGA--TGGCGGCAG--CGGCGCC-TTCCCGCCCGGCCACTTCAAGGACCCCAAG
           200         210         220         230         240         250

       740         750         760         770             780         790         800
    CCGCCAGAGGCCCCCGCAGAGCTCCTCGTCCTGCAGGT-----TGGA-GGCCACGGTGACTGGGTAATCTTG
    | ||   |   |       | |  | || ||||||   |       || ||||  | |||| ||||      |
    CGGCTGTACTGCAAAAACGGGGGCTTCTTCCTGCGCATCCACCCCGACGGCCGAGTTGACGGGGTCCGGGAG
      260         270         280         290         300         310         320

              810         820         830         840         850         860
    AAGCAGGTA--GTCAGA-C-AGCTCACGGATTTTGACAGC-GAAGTCGGAGGAGATGGGGCTGT-GTTGGAA
    ||| ||  |   ||| | | |||| ||  |  || | ||| ||||   |||||| || | || |    |||
    AAG-AGCGACCCTCACATCAAGCT-AC--AACTTCA-AGCAGAAGAGAGAGGAGTTGTGTCTATCAAAGGAG
    330         340         350             360         370         380         390

     870                 880         890         900         910         920
    GGAG------CAGT--CCTGGGTCCCACTG-AGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGATA-GGTTG
     | |      | ||  ||||| |   |  | ||     || | | | |  ||  |    | | | ||  || ||
    TGTGTGCTAACCGTTACCTGGCTATGAAGGAAGATGGAAGATTA-CTG--GCTTCTAAATGTGTTACGGATG
       400         410         420         430         440         450         460

   930         940         950         960             970         980         X
```



```
    TTGGGCTCCAGGCTG-GCGCCAGCACTGTCATTTCGGC------CGGGGGCCCCGGAATTCCGGCCG
       |  ||      ||  ||    | | | | ||  |  |          || |  |||| | |  | |
    AGTGTTTCTTTTTTGAACGATTGGAATCTAATAACTACAATACTTACCGGTCAAGGAAATACACCAGTTGGT
        470         480         490         500         510         520         530

    ATGTG
      540


10. US-08-162-407-5' (1-988)
    Q05883       Plasmid clone pUC-SK1.

 ID   Q05883 standard; DNA; 6767 BP.
 AC   Q05883;
 DT   11-JAN-1991  (first entry)
 DE   Plasmid clone pUC-SK1.
 KW   Basic fibroblast growth factor; tissue regeneration; infarction; ss.
 FH   Key             Location/Qualifiers
 FT   CDS             302..934
 FT   /*tag= a
 FT   /label=basic fibroblast growth factor
 PN   FR2642086-A.
 PD   27-JUL-1990.
 PF   26-JAN-1989; 000973.
 PR   26-JAN-1989; FR-000973.
 PA   (SNFI ) SANOFI SA.
 PI   Caput D, Ferrara P, Kaghad M;
 DR   WPI; 90-277408/37.
 DR   P-PSDB; R06685.
 PT   New recombinant gene encoding basic fibroblast growth factor - in
 PT   new high mol. wt. form, useful e.g. for stimulating tissue
 PT   regenerating or treating infarction
 PS   Disclosure; fig 3; 43pp; French.
 CC   This recombinant gene encodes a basic fibroblast growth factor (bFGF)
 CC   which stimulates growth of mesodermal and neuroectodermal cells.
 CC   It is thus potentially useful e.g. for regenerating damaged tissues,
 CC   and for treating myocardial infarctions, Parkinsons disease and
 CC   Alzheimers disease.  It can be produced on a large scale using rec-
 CC   ombinant DNA methods without risk of contamination.  There are a
 CC   further 3 potential initiation codons allowing expression of larger
 CC   forms of the protein.  See also Q05884.
 SQ   Sequence   6767 BP;    2047 A;    1201 C;    1334 G;    2185 T;

Initial Score    =    137  Optimized Score  =    381  Significance =  8.03
Residue Identity =     51%  Matches          =    442  Mismatches   =   338
Gaps             =     86  Conservative Substitutions               =     0

       170   X   180         190         200         210         220         230
    GCCTCAGTATCCTCCCCTGTAAAATGGGATGA-TAGATGTCTG-TCTCACTGCGTTGTTTAA--GGCTCCGC
              |  |||| | |||   |   || |||  | | |  | ||    |  |   ||  ||   |||
              CGGCCCCAGAAAACCCGAGCGAGTAGGGGGCGGCGCGCAGGAGGGAGGAGAACTGGGGGCGC
              X       10          20          30          40          50          60

         240         250         260         270         280         290         300
    AGGATGAGGCCTTGGCCAG-GTCAGTGCTCCACAAGCAGCAGGTCCTGGGGACTGGGGACGGGGGGCACCTG
     |   |||||  |||   | ||| | | |  | | | | || || |  ||  | |  ||  ||| |||  || |
    GG---GAGGC--TGGTGGGTGTCGGGGGTGGAGATGTAGAAGAT-GTGACGCCGCGGCCCGGCGGGTGCCAG
              70          80          90         100         110         120

          310         320         330         340         350         360         370
```

```
  CTCCCCAGGGCGGGGTGTCCTCCGCCGCGTCCTCTGCCAGTGC-AGGCA-CCAGGCA-GCGGCCAGCAGCAG
   |   | || | || || || | |  ||  |     || | ||   ||| |  || | |||||   |  |||
  ATTAGC-GGAC-GGCTGCCCGCGGTTGCAACGGGATCCCGGGCGCTGCAGCTTGGGAGGCGGCTCTCCCCAG
130         140        150        160        170        180        190

       380        390        400        410        420        430        440
  GAGGCCCACGGGCAG-CAGCAGTAGGAGGAGCAGAGGGGGCTGCGGGGCTGTCGGGGCTGTGGCCTCCAGGG
  | |||   || | || || | |  | | | | |||    | | |||| | |||  |   |  | ||||||
  GCGGCGTCCGCGGAGACACCCATCCGTGAACCCCAGG----TCCCGGGCCGCCGGCTCGCCGCGCACCAGGG
200        210        220        230           240        250        260

       450        460        470        480           490        500
  GCCGG-GGACTCCA-TGGGGGTGGCAGGGTTGAGGAGTCGG----GCTGACACTGCAGCTCCAGGCACCGGG
  ||||| ||||   |  | ||  |   || | ||||    ||    || | | | | || || |   ||   | |
  GCCGGCGGACAGAAGAGCGGCCGAGCGGCTCGAGGCTGGGGGACCGCGGGCGCGGCCGCGCGCTGCCGGGCG
   270        280        290        300        310        320        330

 510        520        530        540        550        560        570
  AGAAGTTCTGGCGAGTGATCCAGGGCTTCAGCGCCAC-CAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAG
   || | |  || |   |  || |||   | | ||| |  ||| | ||||||| |    || || | |||| |
  GGAGGCTGGGGGGCCGGGGCCGGGG---CCGTGCCCCGGAGCGGGTCGGAGG-CCGGGGCCGGGGCCGGGGG
340        350        360           370        380        390        400

580        590        600        610        620        630        640        650
  ATGTTGG-TCTGGACGAAGCGAAGACAGCTGGGGGGGGGCTGAAAGGCACATTTGGTGACAAAGTGTATCTC
  | |  || ||    ||  ||      | ||| ||||    | |   ||| |    || ||| | | | | |
  ACGGCGGCTCCCCGCGCGGCTCCAGCGGCTCGGGGATCCCGGCCGGGCCCCGCAGG-GACCATG-GCAGCCG
   410        420        430        440        450        460        470

          660        670        680        690        700        710        720
  CGTG-TTCA-CGCGCTCCAGCAAGCCTTGCATCTTGGACCCAGCGACAGTCTTGAGCCGCTCCATCCAGCGC
   | |  ||| | ||||   ||  |||||||  |  |||    ||| |||    | |||  |||  || |  |
  GGAGCATCACCACGCT---GCCCGCCTTGC-CCGAGGA--TGGCGGCAG--CGGCGCC-TTCCCGCCCGGCC
   480        490           500        510           520        530

        730           740        750        760        770              780
  TGTGCCAGGACC---AGCCGCCAGAGGCCCCCGCAGAGCTCCTCGTCCTGCAGGT-----TGGA-GGCCACG
    | | ||||||   ||| ||   |   |      | |  | || ||||||   |       || ||||  |
  ACTTCAAGGACCCCAAGCGGCTGTACTGCAAAAACGGGGGCTTCTTCCTGCGCATCCACCCCGACGGCCGAG
540        550        560        570        580        590        600        610

      790        800          810        820        830        840        850
  GTGACTGGGTAATCTTGAAGCAGGTA--GTCAGA-C-AGCTCACGGATTTTGACAGC-GAAGTCGGAGGAGA
   |||| ||||     |||| ||  |   ||| | | |||| ||  |  || | ||| ||||   ||||||
  TTGACGGGGTCCGGGAGAAG-AGCGACCCTCACATCAAGCT-AC--AACTTCA-AGCAGAAGAGAGAGGAGT
        620        630        640        650           660        670

          860        870                880        890        900        910
  TGGGGCTGT-GTTGGAAGGAG------CAGT--CCTGGGTCCCACTG-AGTCCCGAGCTCAGCAGCAGCAGC
  || | || |    |||  | |       | ||  ||||| |   |  | ||      || | | | |  ||  |
  TGTGTCTATCAAAGGAGTGTGTGCTAACCGTTACCTGGCTATGAAGGAAGATGGAAGATTA-CTG--GCTTC
680        690        700        710        720        730        740

      920        930        940        950        960              970
  AGGAGGAGATA-GGTTGTTGGGCTCCAGGCTG-GCGCCAGCACTGTCATTTCGGC------CGGGGGCCCCG
     | | | || || ||    |  ||      ||  ||    | | | | ||  | |           || |  |
  TAAATGTGTTACGGATGAGTGTTTCTTTTTTGAACGATTGGAATCTAATAACTACAATACTTACCGGTCAAG
  750        760        770        780        790        800        810

   980        X
```



```
    GAATTCCGGCCG
    ||| | |  | |
    GAAATACACCAGTTGGTATGTG
  820        830        840
```

11. US-08-162-407-5′ (1-988)
    Q03842 LFA-1 alpha subunit gene

```
ID   Q03842 standard; DNA; 5139 BP.
AC   Q03842;
DT   23-AUG-1990  (first entry)
DE   LFA-1 alpha subunit gene
KW   Lymphocyte function associated antigen; inflammation; metastasis; ss.
FH   Key             Location/Qualifiers
FT   CDS             95..1095
FT   /*tag= a
FT   /label=LFA-1 alpha subunit
FT   sig_peptide     95..67
FT   /*tag= b
FT   misc_RNA        446..490
FT   /*tag= c
FT   /label=tryptic peptide
FT   misc_RNA        770..803
FT   /*tag= d
FT   /label=tryptic peptide
FT   misc_RNA        938..961
FT   /*tag= e
FT   /label=tryptic peptide
FT   misc_RNA        1391..1417
FT   /*tag= f
FT   /label=tryptic peptide
FT   misc_RNA        1655..1687
FT   /*tag= g
FT   /label=tryptic peptide
FT   misc_RNA        1799..1840
FT   /*tag= h
FT   /label=tryptic peptide
FT   misc_RNA        1862..1906
FT   /*tag= i
FT   /label=tryptic peptide
FT   misc_RNA        2585..2626
FT   /*tag= j
FT   /label=tryptic peptide
FT   misc_RNA        2963..3016
FT   /*tag= k
FT   /label=tryptic peptide
FT   misc_feature    3512..3514
FT   /*tag= l
FT   /label=serine phosphorylation site
FT   misc_feature    3527..3529
FT   /*tag= m
FT   /label=serine phosphorylation site
FT   misc_feature    3581..3583
FT   /*tag= n
FT   /label=serine phosphorylation site
FT   misc_feature    3587..3589
FT   /*tag= o
FT   /label=serine phosphorylation site
FT   misc_RNA        3359..3445
```

```
FT   /*tag=  p
FT   /label=transmembrane domain
FT   polyA_signal    5107..5112
FT   /*tag=  q
FT   repeat_region   4623..4912
FT   /*tag=  r
FT   /label=Alu 1
FT   /note="tandem repeat"
FT   misc_feature    287..295
FT   /*tag=  s
FT   /label=N-glycosylation site
FT   misc_feature    359..367
FT   /*tag=  t
FT   /label=N-glycosylation site
FT   misc_feature    656..664
FT   /*tag=  u
FT   /label=N-glycosylation site
FT   misc_feature    2039..2047
FT   /*tag=  v
FT   /label=N-glycosylation site
FT   misc_feature    2102..2110
FT   /*tag=  w
FT   /label=N-glycosylation site
FT   misc_feature    2270..2278
FT   /*tag=  x
FT   /label=N-glycosylation site
FT   misc_feature    2282..2290
FT   /*tag=  y
FT   /label=N-glycosylation site
FT   misc_feature    2678..2686
FT   /*tag=  z
FT   /label=N-glycosylation site
FT   misc_feature    2747..2755
FT   /*tag=  aa
FT   /label=N-glycosylation site
FT   misc_feature    2783..2791
FT   /*tag=  bb
FT   /label=N-glycosylation site
FT   misc_feature    3272..3280
FT   /*tag=  cc
FT   /label=N-glycosylation site
FT   misc_feature    3305..3313
FT   /*tag=  dd
FT   /label=N-glycosylation site
FT   misc_RNA        170..3358
FT   /*tag=  ee
FT   /label=extracellular domain
PN   EP-362526-A.
PD   11-APR-1990.
PF   17-AUG-1989; 115160.
PR   09-MAR-1989; US-321017, US-235227.
PA   (DANA-) Dana Farber Cancer.
PI   Springer TA, Larson R;
DR   WPI; 90-108985/15.
DR   P-PSDB; R05782.
PT   Pure alpha subunit of lymphocyte function associated antigen -
PT   and encoding DNA sequences, useful eg for suppressing
PT   inflammation or metastasis
PS   Disclosure; fig 3; 27pp; English.
CC   This sequence encodes lymphocyte function associated antigen (LFA-1)
```



```
CC   which suppresses inflammation, metastasis and growth of alpha subunit-
CC   expressing tumour cells and is used in the treatment of viral infections.
CC   The pref. dose is 1 pg - 10 mg/kg.
SQ   Sequence   5139 BP;   1165 A;   1439 C;   1399 G;   1135 T;   1 Others;

Initial Score    =    134  Optimized Score  =    378  Significance =  7.83
Residue Identity =    48%  Matches          =    452  Mismatches   =   366
Gaps             =    106  Conservative Substitutions               =     0

  100       110       120       130       140       150       160       170
    GCCTCTAGCCAACTTCCTCTGTGTCCAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTCAGT
                | |||| || | | ||  || | |     | ||  |  | | ||| | |   |||||
              GAATTCC-CTCT-TTCACCCTGTCTA---GGTTGCCAGCAAAT-CCCACGG---GCCTCCTG
              X         10        20        30        40        50

        180       190       200       210       220       230       240
    ATCCTCCCCTGTAAAATGGGATGATAGATGTCTGTCTCACTGCGTTGTTTAAGGCTCCGCAGGATGA-GGCC
    |  || |||       ||||   | || |  ||      ||||   | | |||| | | | | || |  |
    ACGCTGCCC------CTGGGGCCACAGGTCCCTCGAGTGCTGCAAGGATGAAGGATTC-CTGCATCACTGTG
         60              70        80        90       100       110

        250       260       270       280       290       300       310
    TTGGCCAGGTCAGTGCTCCACAAGCAGCAGGTCCTGGGGACTGGGGACGGGGGGC-ACCTGCTC-CCCAGGG
     |||||| | |  ||||   |  |   |   | ||  |  | ||   || |   | |||||  |   | |||
    ATGGCCATGGCGCTGCT-GTCTGGGTTCTTTTTCTTCGCGCCGGCCTCGAGCTACAACCTGGACGTGCGGGG
   120       130       140       150       160       170       180

        320       330       340       350       360       370       380
    CGGGGTGTCCTCCGCCGCGTCCTCTGCCAGTGCAGGCAC-CAGGCAGCGGCCAGCAGCAGGAGGCCCACGGG
    ||  |  || | || |  || | ||     | ||||||   ||   | ||   | |||||  |   ||| |
    CGCGCGGAGCTTCTCC-CCACCGC-GCGCCGGGAGGCACTTTGGATACCGCGTCCTGCAGGTCGGAAACGCG
 190       200       210       220       230       240       250

        390       400       410       420       430       440       450
    CAGCAGCAGTAGGAGGAGCAGAGG-GGGCTGCGGGGC-TGTCGGGGCTGTGGCCTCCAG--GGGCCGGGGAC
      ||   || ||||   ||| || |||   | |  |  |  |   || |     ||||  ||||   ||||
    -GTCA-TTGTGGGAGCTCCAGGGGAGGGGAACAGCACAGGAAGCCTCTATCAGTGCCAGTCGGGCACAGGAC
 260       270       280       290       300       310       320

        460       470       480       490       500       510       520
    TCCATGGGGGTGGCAGGGTTGAGGAGTCGGGCTGACACTGCAGCTCCAGGCACCGGGAGAAGTTC-TGGCGA
     |  ||    || |||       |||   || |   |||  | ||||| | ||  ||  | |  | |||| |
    AC--TG--CCTGCCAGTCACCCTGAG--AGGTTCCAACTATACCTCCAAGTACTTGGGAATGACCTTGGCAA
330           340       350       360       370       380       390

          530       540       550       560       570       580
    GTGATC---CAG--GGCTTCAGCGCCACCAGCTG---CTCGGAGGTCTCCTGCAGGAGGCGGG---AGATGT
      || |   |||  ||   ||       || | ||   || ||  |||||   |  | | | |   |  | | |
    CAGACCCCACAGATGGAAGCATTTTGGCCTG-TGACCCTGGGCTGTCTCGAACGTGTGACCAGAACACCTAT
      400       410       420       430       440       450       460

        590       600       610       620       630       640       650
    TGGTCTGGACGAAGCGAAGACAGCTGGGGGGGGGCTGAAAGG-CACATTTGGTGACAAAGTGTATCTCCGTG
      |  ||| |   |  |   |  | |  |   |  ||| | || | ||  || || ||  | |   |   ||
    CTGAGTGGCCTGTGTTACCTCTTC-CGCCAGAATCTGCAGGGTCCCA--TGCTG-CA-GGGGCGCCCTGGTT
     470       480       490       500       510       520       530

       660       670       680       690       700       710       720
    TTCACGCGCTCCAGCAAGCCTTGCATCTTGGACCCAGCGACAGTCTTGAGCCGCTCCATCCAGCGCTGTGCC
    |||| |   |  | ||||    ||| | | ||||       || ||   | | |  ||    || || ||
```

```
 470       480       490       500       510       520       530
GGGTTGAGGAGTCGGGCTGACACTGCAGCTCCAGGCACCG-GGAGAAGTTCTGGCGAG-TG--ATCCAGGGC
      |||||    ||||  | ||    | |  || ||| | | |  || || |  ||| ||  | |||| |
TCCCCAAGGAG-ATGGCT-TCTCTCTCGGTGGAGACACAGAGCACCAGCTCGGAGGAGATGGTACCCAGCTC
  480       490       500       510       520       530       540

      540       550       560       570       580       590       600
T---TCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATG-TTGGTCTGGACGAAGCGAAG
|   ||| | ||||||  || ||||  |||||      |  || |      ||| || || ||  | | |  |||
TCCCTCACCCCCACCACCTCCTCG--GGTCT-----ATAAGCC-----ATGCTTTGTATG--CAATGACAAG
  550       560       570             580             590       600

     610       620       630       640       650       660       670
ACAGCTGGGGGGGGGCTGAAAGGCACATTTGGTGACAAAG-TGTATCTCCGTGTTCACGCGCTCCAGCAAGC
 |  || ||       || |  ||  ||   |  ||  || | || |     | | ||||  |||  ||   ||
TCTTCT-GGCTACCACT-ATGGGGTCAGCTCCTGTGAAGGCTGCAAGGGCTTCTTCAGACGCAGCATTCAGA
     610       620       630       640       650       660       670

    680       690           700       710       720       730       740
CTTGCATCTTGGACCCA-GCGACAGT--CTTGAGCCGCTCCATCCAGCGCTGT-GCCAGGACCAGCCGCCAG
    |||  || |  || |  || ||   |   | | |   ||| || |   ||  |||| |   |  |||||
AAAACATGGTGTATACATGTCACCGTGACAAAAACTGTATCAT-CAACAAGGTCACCAGAAATCGATGCCAG
    680       690       700       710       720       730       740

      750         760         770         780       790         800
AGGCCCCCGCAGAGCTCC--TCGTCCTGC-AGGTTGG-AGG-CCACGGTGACTG---GGTAATCTTGAAGCA
       | |||| ||| |    ||  | | | || || | | ||| ||   |||  || |   | ||| ||
----TACTGCAG-GCTACAAAAGTGTTTCGAAGTGGGCATGTCCAAGGAAGCTGTAAGGAACGATCGAAACA
       750         760         770         780       790         800

  810       820       830       840       850       860       870
GGTAGTCAGA-CAGCTCACGGATTTTGACAGCGAAGTCGG--AGGAGATG-GGGCTGTGTTGGAAGGA-GCA
 | ||  | |  || | |  ||    | |       | | |  | ||  || | |   |  ||| || | |  ||
AGAAGAAAAAGGAGGTAAAAGAGGAGGGCTCGCCCGACAGCTATGAACTGAGTCCACAGTTAGAGGAACTCA
      820       830       840       850       860       870       880

      880       890             900       910       920       930
-GTCCTGGGTC-CCACTGAGTCCC--GAG-CT----CAGCA-GCAG-CAGCAGGAGGAGATAGGTTGTTGGG
   ||  ||||  ||  |    ||  ||| ||     |  ||  | | |||| || | |  ||        |
TCACCAAGGTCAGCAAAGCCCACCAGGAGACTTTTCCCTCACTCTGCCAGCTGG-GCAAGTA-CACCACGAA
    890       900       910       920       930       940       950

     940       950       960       970           980          X
CTCCAG-GCTG-GCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGG----AATTCCG---GCCG
|||||| || |  | || |   || ||  | | || |||||   ||    | ||| |   || |
CTCCAGTGCAGATCACCGG--GTG-CAGCTGGACCTGGGGCTGTGGGACAAGTTCAGCGAGCTGGCCACCAA
    960       970       980       990      1000      1010   X  1020

AT
```

13. US-08-162-407-5' (1-988)
    Q20541 Mouse retinoic acid receptor isoform clone lambda-

ID Q20541 standard; DNA; 2740 BP.
AC Q20541;
DT 10-MAR-1993 (revised)
DT 22-APR-1992 (first entry)

DE Mouse retinoic acid receptor isoform clone lambda-mE13.20.
KW RAR-gamma-A; ss.
OS Mus musculus.

FH	Key	Location/Qualifiers
FT	CDS	253..1629
FT	/*tag= a	
FT	polyA_signal	2697..2702
FT	/*tag= b	
FT	polyA_site	2716..2740
FT	/*tag= c	
FT	misc_feature	1..435
FT	/*tag= d	
FT	/note= "point of divergence from RAR-gamma-B"	

PN CA2015766-A.
PD 31-OCT-1991.
PF 30-APR-1990; 015766.
PR 30-APR-1990; CA-015766.
PA (MOUN) MOUNT SINAI HOSPITAL.
PA (HSCR-) HSC RES DEV CORP.
PI Giguere V, Varmuza SL, Rossant J;
DR WPI; 92-024769/04.
DR P-PSDB; R20465.
PT DNA segments to produce iso-forms and chimeric forms of RAR -
PT have sequence encoding protein which display ligand binding and
PT transcription activating properties of a gamma iso-form of RAR
PS Claim 5; Fig 1; 52pp; English.
CC The sequence is that of a clone lambda-mE13.20 which encodes mouse
CC retinoic acid receptor (RAR) isoform RAR-gamma-A, it was isolated
CC from a 7.5 day mouse embryo lambda-gt10 cDNA library. It diverges
CC from the clone lambda-mE7.1 as indicated at nucleotides 1 to 435,
CC this corresponds to with a splice junction between exon 1 and exon
CC 2 of the human RAR-alpha gene. It can be used for screening of
CC agonists and antagonists of the gamma receptor effector system and
CC for identifying ligands capable of binding to a gamma isoform of RAR.
CC See also Q20542.
SQ Sequence 2740 BP; 602 A; 840 C; 741 G; 557 T;

Initial Score	= 132	Optimized Score	= 417	Significance	= 7.70
Residue Identity	= 50%	Matches	= 508	Mismatches	= 380
Gaps	= 125	Conservative Substitutions		=	0

```
    30        40          50        60        70        80        90
    GGGGCCGGTGCTGGATTCT--GGCGTTGCCTCCATTCTGGGGAGGGAATGAGAGGGGCCCAAGGAAGGGACC
              | ||  |||  ||| || | |||   | |           ||  |   |||    |||  ||
              CGGGGCTCTGCGGCCTTCCTTCC--CCCGCCCTCCCCTCCAGCAGTTTCCA--CCAGGTCCC
              X       10        20        30        40        50

    100       110       120       130       140       150       160
    -GGCCTCTAGCCAACTTCCTCTGTGT-CCAGGCTATACATCCTCTGGCTGGTGACTCCCCTCTGTGTGCCTC
       |||| ||||      ||  | |||   |||  ||   | |   |||  | |     |||  |    || |
    TCACCTC-AGCC-TGGCCCAGTATGTAGGAGGGAATCTCTGCAGAGGCCAGAG-GGATCCTTGGAAACCCAC
    60          70        80        90        100       110       120

   170       180       190       200       210       220       230
    AGTATCCTCCCCTGTAAAATGGG-ATGATAGATGTCTGTCTCACTGCGTTGTTTAAGGCTCCGCAGGATGAG
     | | |   ||  | |   |||| | |      |  | | | ||  ||  ||   | ||   ||  || ||
    TGGA-CAGACCAGGCAGGGTGGGCACGGAGCCTCCCAGGC-CAGGGCAGTGGGCATGG--GCGGGGGCTGTA
    130       140       150       160       170       180       190

  240       250       260       270       280       290       300
```

```
GCCTTGGCCAGGTCAGTGCTCCACAAGCAGCAG-GTCCTGGGGACTGGGGACG-GGGGGC-ACC-TGCTCCC
||  ||   ||  ||   | || |  || |||| |||| |||| |    |||  |  || ||| ||  | |
GC--TG--AAGACCA--CCCCCGCCTGCTGCAGAGTCCAAGGGA-TTCCCACGCCGCAGCTACCATGGCCAC
      200       210       220       230       240       250       260

310       320       330       340       350       360       370
CAGGGCGGGGTGTC-CT--CCGCCGCGTCCTCTGCCAGTGCAGGCACCAGGCAGCGGCCAGCAGCAGGAGGC
||    || | | | ||   |||| || | |||   | ||  | | | ||   |  |||| ||||||   |
CAATAAGGAGAGACTCTTTGCGCCCGGTGCCCTG---GGGCCTGGATCTGGTTAC--CCAGGAGCAGGCTTC
       270       280       290       300       310       320

 380       390       400       410       420       430       440
CCACGGGCAGCAGCAGTAGGAGGAGCAGAGGG-GGCTGC----GGGGCTGTCGGGGCTGTGGCCTCCAGGGG
|||   ||     |||  | |    |||||||  ||  |     | | || | | ||  | || ||| ||||
CCATTCGCCTTCCCAGGTGCA--CTCAGAGGGTCGCCACCATTTGAGATG-CTGAGCCCTAGCTTCC-GGGG
330       340       350       360       370       380       390

   450       460       470       480       490       500       510
CCGGGGACTCCATGGGGGTGGCAGGGTTGAGGAGTCGGGCTGACACTGCAGCTCCAGGCACCG-GGAGAAGT
|| |||   |||   |  || |       |||||   ||||  | ||   | |  || ||| | | |  ||
CCTGGG---CCA---GCCTGACCTCCCCAAGGAG-ATGGCT-TCTCTCTCGGTGGAGACACAGAGCACCAGC
  400               410       420       430       440       450

   520           530           540       550       560       570       580
TCTGGCGAG-TG--ATCCAGGGCT---TCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGA
|| |  ||| ||  | ||||  ||   ||| | |||||| || ||||  |||||      |  || |     |
TCGGAGGAGATGGTACCCAGCTCTCCCTCACCCCCACCACCTCCTCG--GGTCT-----ATAAGCC-----A
460       470       480       490       500       510

         590       600       610       620       630       640       650
TG-TTGGTCTGGACGAAGCGAAGACAGCTGGGGGGGGGCTGAAAGGCACATTTGGTGACAAAG-TGTATCTC
|| || || ||  | | |  ||| |  || ||       || |  ||  ||  |  ||  || | || |
TGCTTTGTATG--CAATGACAAGTCTTCT-GGCTACCACT-ATGGGGTCAGCTCCTGTGAAGGCTGCAAGGG
520       530       540       550       560       570       580

         660       670       680       690       700       710
CGTGTTCACGCGCTCCAGCAAGCCTTGCATCTTGGACCCA-GCGACAGT--CTTGAGCCGCTCCATCCAGCG
| | ||||  |||  ||   ||    |||  || |  || |  || ||  |   | | |   ||| || |
CTTCTTCAGACGCAGCATTCAGAAAAACATGGTGTATACATGTCACCGTGACAAAAACTGTATCAT-CAACA
590       600       610       620       630       640       650

720       730       740       750       760       770       780
CTGT-GCCAGGACCAGCCGCCAGAGGCCCCCGCAGAGCTCC--TCGTCCTGC-AGGTTGG-AGG-CCACGGT
   ||  |||| |   |  |||||       | |||| ||| |    ||  | | | || || | | ||| ||
AGGTCACCAGAAATCGATGCCAG----TACTGCAG-GCTACAAAAGTGTTTCGAAGTGGGCATGTCCAAGGA
660       670       680           690       700       710       720

  790           800       810       820       830       840       850
GACTG---GGTAATCTTGAAGCAGGTAGTCAGA-CAGCTCACGGATTTTGACAGCGAAGTCGG--AGGAGAT
  |||   || |   | ||| || | ||  | |  || | |  ||    | |         | | |  | ||  |
AGCTGTAAGGAACGATCGAAACAAGAAGAAAAAGGAGGTAAAAGAGGAGGGCTCGCCCGACAGCTATGAACT
  730       740       750       760       770       780       790

        860       870       880       890           900           910
G-GGGCTGTGTTGGAAGGA-GCA-GTCCTGGGTC-CCACTGAGTCCC--GAG-CT----CAGCA-GCAG-CA
| |  |  ||| || | |  ||   ||  |||| ||  |    ||  ||| ||     |  ||  | | ||
GAGTCCACAGTTAGAGGAACTCATCACCAAGGTCAGCAAAGCCCACCAGGAGACTTTTCCCTCACTCTGCCA
800       810       820       830       840       850       860

        920       930       940       950       960       970
```



```
GCAGGAGGAGATAGGTTGTTGGGCTCCAG-GCTG-GCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGG---
|| || | |  ||       |  |||||| || |  | || |   || ||  | | || |||||   ||
GCTGG-GCAAGTA-CACCACGAACTCCAGTGCAGATCACCGG--GTG-CAGCTGGACCTGGGGCTGTGGGAC
870       880       890       900       910       920       930

 980          X
-AATTCCG---GCCG
 | ||| |   || |
AAGTTCAGCGAGCTGGCCACCAAAT
 940       950       960
```

14. US-08-162-407-5' (1-988)
    Q14498 Clone 1 for truncated poly Ig-receptor.

```
ID   Q14498 standard; DNA; 1876 BP.
AC   Q14498;
DT   10-MAR-1993  (revised)
DT   30-JAN-1992  (first entry)
DE   Clone 1 for truncated poly Ig-receptor.
KW   Rabbit; insemination; pregnancy; ss.
OS   Oryctolagus cuniculus.
FH   Key             Location/Qualifiers
FT   CDS             124..1876
PN   WO9116061-A.
PD   31-OCT-1991.
PF   16-APR-1991; U02604.
PR   16-APR-1990; US-510161.
PA   (HARD ) HARVARD COLLEGE.
PA   (SURE-) INST SUISSE RECH EXPER C.
PI   Kraehenbuhl JP, Weltzin RA, Neutra MR;
DR   WPI; 91-339549/46.
DR   P-PSDB; R14670.
PT   Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
PT   useful in protection against pathogens or against pregnancy
PS   Disclosure; Fig 8; 51 pp; English.
CC   The sequence was obtd. from a cDNA clone of alllele no. 1 and
CC   encodes a truncated poly-Ig receptor.  The native gene (Mostov et
CC   al) is mutated to delete the portion encoding the transmembrane and
CC   intra-cellular domains.  The recombinant protein produced by
CC   expression of the sequence is used as a stabiliser protein with a
CC   poly-Ig specific for a selected antigen or family of antigens. The
CC   compsn. can be administered directly to the mucosal surfaces of a
CC   mammal to protect against a pathogen or against insemination. It
CC   protects against allergens that contact the respiratory or digestive
CC   mucosal surfaces and protects against pregnancy by cross-linking
CC   sperm in the vagina.
CC   See also Q14499.
SQ   Sequence   1876 BP;    441 A;    525 C;    549 G;    361 T;

Initial Score    =    130  Optimized Score  =    263  Significance =  7.56
Residue Identity =    51%  Matches         =    325  Mismatches   =   226
Gaps             =     85  Conservative Substitutions              =     0

        400  X    410       420       430       440       450       460
   TAGGAGGAGCAGAGGGGGCTGCGGGGCTGTCGG--GGCTGTGGCCTCCAGGGGCCGGGGACTCCATG-GGGG
              |  |||| | ||||   | | |  |||||||  | || | | ||  || |  || | ||||
            GGCCGGGGTTACGGGCTGGCCAGCAGGCTGTG--CCCCCGAGTCC--GGTCAGCAGGAGGGG
            X       10        20        30          40          50
```

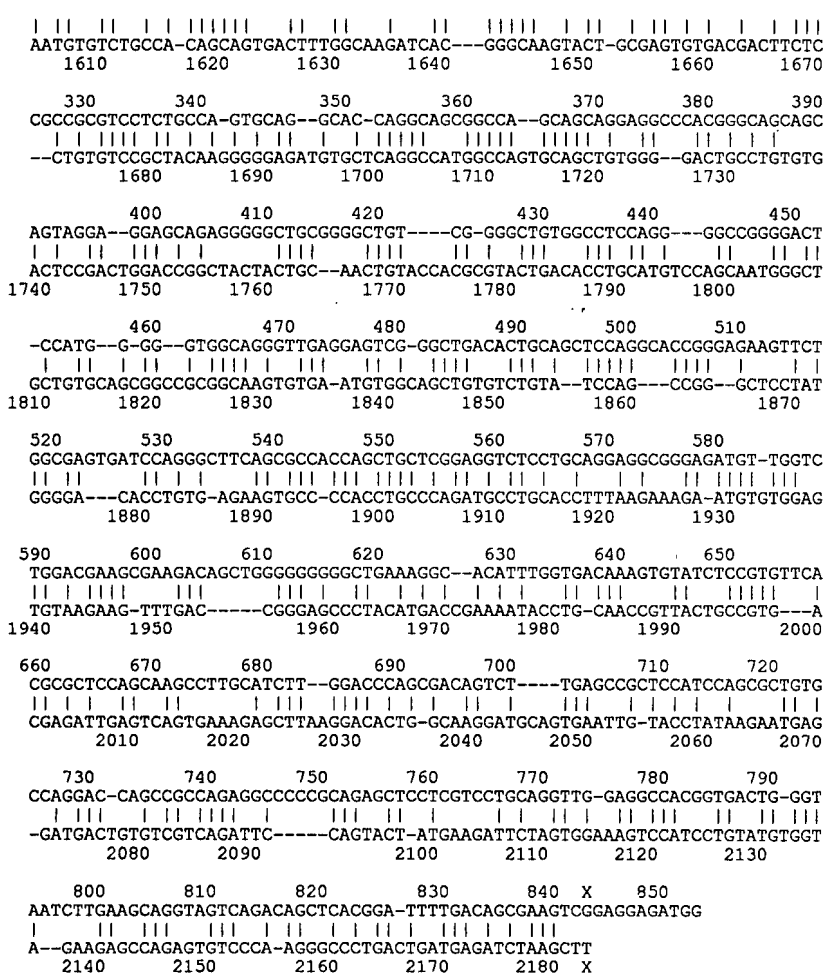
```
      470       480       490       500       510       520       530
TGGCAGGGTTGAGGAGTCGGGCTGACA-CTGCAGCTCCAGGCACCGGGAGAAGTTCTGGCGAGTGATCCAGG
  | || |       | ||   |   || | ||||| |||||  ||  | |   | |  || |   | ||| |
AAGAAGTGGCCTAAAATCTCTCCCGCATCGGCAGC-CCAGG--CCTAGTGCCCTACCAGCCA-CCAGCCATG
60        70        80        90        100       110       120

      540           550       560       570           580       590
GCT-TCAGC--GC-CACCAG-CTGCTCGGAGGTCTCCTGCA--GGAGGCGG----GAGATGTTGGTCTGGAC
||| ||  |  || |||| | ||||| ||  |||| | ||  ||   |||     || |  |  |  || |
GCTCTCTTCTTGCTCACCTGCCTGCT-GGCTGTCT-TTTCAGCGGCCACGGCACAAAGCTCCTTATTGGGTC
 130       140       150       160       170       180       190

    600       610           620       630       640       650       660
GAAGCGAAGACAGCTGG---GGGGGGGCTGAAAGGCACATTTGGTGACAAAGTGTATCTC-CGTGTTCACGC
  |||     | |  |||    |||| | |||| |       || |||   ||||   || | || | ||
CCAGC-TCCATATTTGGTCCCGGGGAGGTGAATGTTTTGGAAGGCGACTCGGTGTCCATCACATGCT-AC-T
 200       210       220       230       240       250       260

      670       680       690       700           710       720       730
GCTCCAGCAAGCCTTGCATCTTGGACCCAGCGACAGTCTTGAG--CCGCT-CCATCCAGCGCTGTGCCAGGA
 | ||| ||| |||  | ||    |||| || |||| |   ||  | | | ||    || |  | ||  ||
AC-CCAACAA-CCT-CCGTC----ACCCGGC-ACAGCCGGAAGTTCTGGTGCCGGGAAGAGGAGAGC--GG-
   270       280           290       300       310       320

       740           750       760       770               780       790
CCAGCCGCCAGAGGC---CCCCGCAGAGCTCCTCGTCCTGCAGG-------TTGGAG-GCCACGGTGACTGG
 | || ||  || ||   || ||    ||| | |||||  ||||          |||| | || | | || |
-CCGCTGCGTGACGCTTGCCTCGACCGGCTACACGTCC--CAGGAATACTCCGGGAGAGGCAAGCTCACCGA
 330       340       350       360       370       380       390

        800       810       820       830       840       850
GTAATCT--TGAA--GCAGGTAGT-CAGACAGCTCACGGATTTTGACAG--CGAAGTCGGAGGAGATGGGGC
 |   ||  | ||  | || | ||   ||| | | ||  | |    |||  || | || | |||| |
CTTCCCTGATAAAGGGGAGTTTGTGGTGACTGTTGACCAACTCACCCAGAACG-ACTCAG-GGAGCTACAAG
 400       410       420       430       440       450       460

 860           870       880       890       900           910       920
TGTG--TTGGAAG-GAGCAGTCCTGGGTCCCACTGAGTCCCGAGCTCAGC-AGCAG--CAGCAGGAGGAGAT
||||   ||| ||  | | | | ||| ||   || | | | | ||| |  || |  |||| || |  |
TGTGGCGTGGGAGTCAACGGCCGTGG--CC--TGGACTTCGGTG-TCAACGTGCTGGTCAGCCAGAAGCCAG
 470       480       490           500       510       520       530

     930       940       950       960       970       980         X
AGGTTGTTG-GGCTCCAGGCTGGCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGG--AATTC-CGGCCG
||  || ||  | |      |   |   |  | || | | | |       | || |    ||| |    |
AGCCTGATGACGTTGTTTACAAACAATATGAGAGTTA-TACAGTAACCATCACCTGCCCTTTCACATATGCG
     540       550       560       570       580       590       600

ACTAGGCA
    610
```

15. US-08-162-407-5' (1-988)
    Q05272 Sequence encoding secreted GPIIIa subunit of multi

ID Q05272 standard; DNA; 2183 BP.
AC Q05272;
DT 22-NOV-1990 (first entry)
DE Sequence encoding secreted GPIIIa subunit of multiple subunit
DE polypeptide (MSP) GPIIb-IIIa.

KW GPIIb; GPIIIa; MSP GPIIb-IIIa; anti-coagulant; anti-inflamatory;
KW immunosuppressant; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 13..2166
FT /*tag= a
FT misc_signal 89
FT /*tag= b
FT /label=Signal processing site.
PN WO9006953-A.
PD 28-JUN-1990.
PF 20-DEC-1989; U05743.
PR 22-DEC-1988; US-290224.
PR 01-DEC-1989; US-444490.
PA (GETH) GENENTECH INC.
PI BODARY SC, GORMAN CM, NAPIER MA, MCLEAN JW;
DR WPI; 90-224496/29.
DR P-PSDB; R05936.
PT Soluble analogues of multi-sub-unit polypeptide prodn. - by
PT transforming host cells with nucleic acid modified to prevent
PT membrane attachment, useful in diagnosis and ligand purificn.
PS Disclosure; 47pp; English.
CC Peptide product is analogous to MSP with better stability, it is
CC not membrane bound and so may be collected as a cell secretion from
CC a transformed host. The product is useful as a diagnostic reagent
CC eg. immunoassay of MSP; in purification of ligands and matrix
CC proteins; and therapeutically as agonists or antagonists of the
CC corresponding membrane bound receptor.
SQ Sequence 2183 BP; 506 A; 563 C; 621 G; 493 T;

```
Initial Score    =    126  Optimized Score  =    368  Significance =  7.29
Residue Identity =    50%  Matches         =    455  Mismatches   =   323
Gaps             =    127  Conservative Substitutions             =     0

             X       10        20        30        40        50
             CGGGGGCTTTGTTCAGAGTTGGGTAAATGGGGCCGGT--GCT-GGATTCTGGCGTTGCCT-C
             || ||||    ||||  |   |    || | |  | |||   ||  |||| || | |||| |
    CCATAAAGCCCGTGGGC----TTCAAGGACAGCCTGATCGTCCAGGTCACCTTTGATTGTGACTGTGCCTGC
       1340   X 1350          1360      1370      1380      1390      1400

    60        70         80        90        100       110       120
    CATTCTGGGGAGGGA-ATGAGAGGGGCCCAAGGAAGGGAC-CGGCCTCTAGCCAACTTCCTCTGTGTCCAGG
    ||  |    |  | |  | | ||    |||  |  |  ||   | |  | |  | |||  |  ||||   ||
    CAGGCCCAAGCTGAACCTAATAG----CCATCGCTGCAACAATGGCAAT-GGGACCTT--TGAGTGT--GGG
         1410      1420          1430      1440      1450      1460

   130           140       150       160           170       180
    CTAT-ACATCCT----CTGGCTGG-TGACTCCCCTCTGTGTG-CCTC-----AGTATCCTC-CCCTGTAAAA
     |||  | |  |    |||||||| ||    |||  |||| |  |||      || |   || ||||    |
    GTATGCCGTTGTGGGCCTGGCTGGCTGGGATCCCAGTGTGAGTGCTCAGAGGAGGACTATCGCCCTTCCCAG
      1470      1480      1490      1500      1510      1520      1530

    190       200       210           220       230       240       250
    TGGGATGATAGATGTCTGTCTCACTGCGT-----TGTTTAAGGCTCCGCAGGATGAG-GCCTTGGCCAGGTC
      ||| ||  |  | |   |  |  | ||     || |   || | || ||  ||| ||||  |   ||||
    CAGGACGAGTGCAGCC--CCCGAGAGGGTCAGCCCGTCTGCAGC-CAGCGGGGCGAGTGCCTCTG--TGGTC
    1540      1550          1560      1570      1580      1590      1600

        260       270       280       290       300       310       320
    AGTG-CTCCACAAGCAGCAGGTCCTGGGGACTGGGGACGGGGGGCACCTGCTCCCCAGGGCGGGGTGTCCTC
```


AATGTGTCTGCCA-CAGCAGTGACTTTGGCAAGATCAC---GGGCAAGTACT-GCGAGTGTGACGACTTCTC
1610 1620 1630 1640 1650 1660 1670
330 340 350 360 370 380 390
CGCCGCGTCCTCTGCCA-GTGCAG--GCAC-CAGGCAGCGGCCA--GCAGCAGGAGGCCCACGGGCAGCAGC
--CTGTGTCCGCTACAAGGGGGAGATGTGCTCAGGCCATGGCCAGTGCAGCTGTGGG--GACTGCCTGTGTG
1680 1690 1700 1710 1720 1730
400 410 420 430 440 450
AGTAGGA--GGAGCAGAGGGGGCTGCGGGGCTGT----CG-GGGCTGTGGCCTCCAGG---GGCCGGGGACT
ACTCCGACTGGACCGGCTACTACTGC--AACTGTACCACGCGTACTGACACCTGCATGTCCAGCAATGGGCT
1740 1750 1760 1770 1780 1790 1800
460 470 480 490 500 510
-CCATG--G-GG--GTGGCAGGGTTGAGGAGTCG-GGCTGACACTGCAGCTCCAGGCACCGGGAGAAGTTCT
GCTGTGCAGCGGCCGCGGCAAGTGTGA-ATGTGGCAGCTGTGTCTGTA--TCCAG---CCGG--GCTCCTAT
1810 1820 1830 1840 1850 1860 1870
520 530 540 550 560 570 580
GGCGAGTGATCCAGGGCTTCAGCGCCACCAGCTGCTCGGAGGTCTCCTGCAGGAGGCGGGAGATGT-TGGTC
GGGGA---CACCTGTG-AGAAGTGCC-CCACCTGCCCAGATGCCTGCACCTTTAAGAAAGA-ATGTGTGGAG
1880 1890 1900 1910 1920 1930
590 600 610 620 630 640 650
TGGACGAAGCGAAGACAGCTGGGGGGGGGCTGAAAGGC--ACATTTGGTGACAAAGTGTATCTCCGTGTTCA
TGTAAGAAG-TTTGAC-----CGGGAGCCCTACATGACCGAAAATACCTG-CAACCGTTACTGCCGTG---A
1940 1950 1960 1970 1980 1990 2000
660 670 680 690 700 710 720
CGCGCTCCAGCAAGCCTTGCATCTT--GGACCCAGCGACAGTCT----TGAGCCGCTCCATCCAGCGCTGTG
CGAGATTGAGTCAGTGAAAGAGCTTAAGGACACTG-GCAAGGATGCAGTGAATTG-TACCTATAAGAATGAG
2010 2020 2030 2040 2050 2060 2070
730 740 750 760 770 780 790
CCAGGAC-CAGCCGCCAGAGGCCCCCGCAGAGCTCCTCGTCCTGCAGGTTG-GAGGCCACGGTGACTG-GGT
-GATGACTGTGTCGTCAGATTC-----CAGTACT-ATGAAGATTCTAGTGGAAAGTCCATCCTGTATGTGGT
2080 2090 2100 2110 2120 2130
800 810 820 830 840 X 850
AATCTTGAAGCAGGTAGTCAGACAGCTCACGGA-TTTTGACAGCGAAGTCGGAGGAGATGG
A--GAAGAGCCAGAGTGTCCCA-AGGGCCCTGACTGATGAGATCTAAGCTT
2140 2150 2160 2170 2180 X

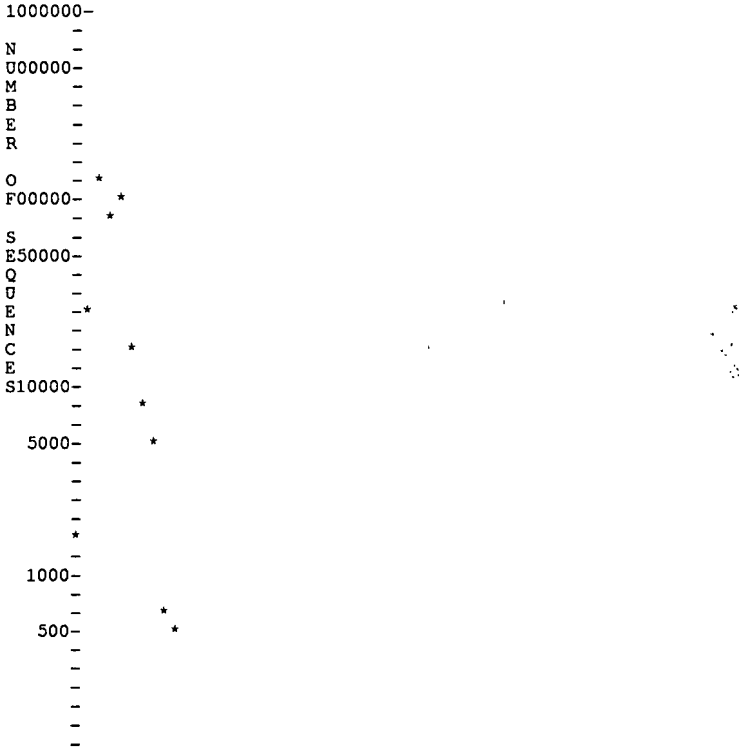
NeWSprinter20

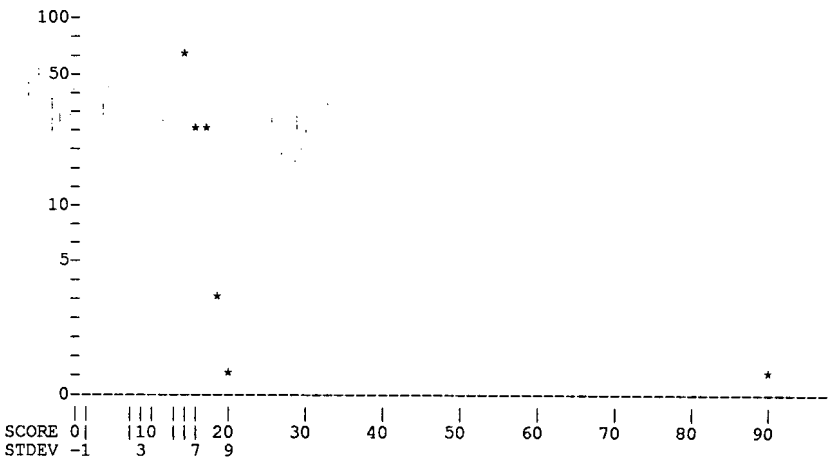
Thu Apr 06 10:17:12 1995

NeWSprint 2.5 Rev B
Openwin library 3
NeWSprint interpreter 210.0

NeWSprint 2.5

```
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq6_28na.res made by  on Wed 5 Apr 95 17:26:04-PDT.


Query sequence being compared:US-08-162-407-6 (1-235)
Number of sequences searched:                313646
Number of scores above cutoff:                 3849

      Results of the initial comparison of US-08-162-407-6 (1-235) with:
   Data bank : EMBL-NEW 1, all entries
   Data bank : GenBank 86, all entries
   Data bank : GenBank-NEW 1, all entries
   Data bank : UEMBL 41_86, all entries
```





PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Translation Frame	1		
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	1		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	4	6	1.79

Times:	CPU	Total Elapsed
	00:14:41.08	00:15:02.00

Number of residues: 96133774
Number of sequences searched: 313646
Number of scores above cutoff: 3849

Cut-off raised to 4.
Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.
Cut-off raised to 11.
Cut-off raised to 12.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 48 standard deviations above mean ****					
1. MMU04807	Mus musculus FLT3/FLK2 ligand	384	90	111	48.15	1
	**** 9 standard deviations above mean ****					
2. SSVHNF1	S.scrofa mRNA for vHNF-1.	940	21	25	9.52	1
	**** 8 standard deviations above mean ****					
3. AIBDVVP2	Avian infectious bursal disea	490	19	26	8.40	1
4. IBDLGVP2	Infectious bursal disease vir	502	19	26	8.40	1
5. SCINVE	Yeast gene SUC2 encoding inve	957	19	32	8.40	1
	**** 7 standard deviations above mean ****					
6. ATT13408	1474 Arabidopsis thaliana cDN	124	18	26	7.84	1
7. T04557	604 Arabidopsis thaliana cDNA	124	18	26	7.84	1
8. SFIURA5A	Sordaria macrospora orotate p	415	18	31	7.84	1
9. RATNOSI	Rat nitric oxide synthase mRN	1198	18	35	7.84	1
10. RATNWUT	Rat mRNA.	1204	18	35	7.84	1
11. MUSINOSL	Mus musculus nitric oxide syn	1230	18	34	7.84	1
12. RNU03699	Rattus norvegicus Sprague Daw	1314	18	35	7.84	1
13. MUSRGEETS	Mouse 18S ribosomal RNA gene	1398	18	30	7.84	1
14. T35584	EST87799 Homo sapiens cDNA 5'	106	17	22	7.28	1
15. CHKRERYF1	Chicken erythroid-specific tr	356	17	31	7.28	1
16. CHKHOX7	Chicken mRNA for Chox-7 prote	442	17	29	7.28	1
17. SYORPOD1	Synechococcus sp. rpoD1 gene	453	17	32	7.28	1
18. AFPECB	Aspergillus flavus 70 polygal	467	17	36	7.28	1
19. YSKHAP3A	Kluyveromyces lactis (HAP3) g	833	17	26	7.28	1
20. MUSTWISTG	Mouse M-twist gene, complete	920	17	31	7.28	1
21. SMIGCU	Suncus murinus mu-immunoglobu	947	17	28	7.28	1
22. RMU10426	Rhodothermus marinus R-21 pla	978	17	28	7.28	1
23. SCSUC1G	Yeast SUC1 gene (EC 3.2.1.26)	1000	17	30	7.28	1
24. BOVNE3B	Bos taurus neurexin III-beta	1135	17	32	7.28	1
25. HUMCACH1A	Human neuroendocrine/beta-cel	2397	17	30	7.28	1
	**** 6 standard deviations above mean ****					
26. MUS68E11	Mouse embryonal carcinoma F9	86	16	20	6.72	1
27. MM68E11	Mouse embryonal carcinoma F9	86	16	20	6.72	1
28. T38268	EST103700 Saccharomyces cerev	128	16	23	6.72	1
29. MMLRP	M.musculus Lrp, alpha-2-Mr ge	233	16	30	6.72	1
30. S66868S1	platelet-derived growth facto	466	16	32	6.72	1
31. DROTFIIB	Drosophila melanogaster RNA p	488	16	33	6.72	1
32. LESALMSGB	L.esculentum S-adenosyl-L-met	555	16	29	6.72	1
33. TMALPTUB	T.marmorata mRNA for alpha-tu	566	16	33	6.72	1
34. MUSAPRF	Mus musculus acute phase resp	884	16	31	6.72	1
35. MMAPRF	Mus musculus acute phase resp	884	16	31	6.72	1
36. CHKHSF3C	Chicken mRNA sequence.	891	16	31	6.72	1
37. CEPMTTGRG	Cephalophus maxwelli mitochon	893	16	31	6.72	1
38. OCPHOR	Rabbit mRNA for muscle glycog	943	16	29	6.72	1
39. RABPYGM	Rabbit muscle glycogen phosph	945	16	29	6.72	1
40. CHKMYOSIND	Chicken fast myosin heavy cha	981	16	29	6.72	1
41. OPOHBBE	Opossum beta-hemoglobin epsil	1194	16	28	6.72	1
42. DVHB	Drosophila virilis hunchback	2935	16	30	6.72	1
43. RATCYP45I	Rat cytochrome P450IIE1 gene,	4242	16	31	6.72	1

44. HSAPOBA2	Human apolipoprotein B-100 mR	4624	16	32	6.72	1
45. CHKMYHE	Chicken embryonic myosin heav	10370	16	29	6.72	1

1. US-08-162-407-6 (1-235)
   MMU04807 Mus musculus FLT3/FLK2 ligand mRNA, complete cds.

```
LOCUS       MMU04807     1152 bp    mRNA            ROD       11-MAY-1994
DEFINITION  Mus musculus FLT3/FLK2 ligand mRNA, complete cds.
ACCESSION   U04807
KEYWORDS    .
SOURCE      mouse.
  ORGANISM  Mus musculus
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Rodentia; Myomorpha; Muridae; Mus.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
            Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
            Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
            Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
            Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
  TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
            haematopoietic stem cells and is encoded by variant RNAs
  JOURNAL   Nature 368, 643-648 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1152)
  AUTHORS   Culpepper,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
            Institute, 901 California Avenue, Palo Alto, CA 94304, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 483846
FEATURES             Location/Qualifiers
     source          1..1152
                     /clone="mouse Flt3/Flk2 ligand T110"
                     /clone_lib="TA4 cDNA"
                     /organism="Mus musculus"
                     /cell_line="TA4 stroma"
     CDS             256..954
                     /note="NCBI gi: 483847"
                     /codon_start=1
                     /product="FLT3/FLK2 ligand"
                     /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                     KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                     NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLGLKPCIGKACQNFSRCLEVQ
                     CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLLPLTLVLLAAAWGLRWQRARR
                     RGELHPGVPLPSHP"
BASE COUNT      238 a    346 c    324 g    244 t
ORIGIN
```

```
Initial Score    =     90  Optimized Score  =    111  Significance = 48.15
Residue Identity =    71%  Matches          =    112  Mismatches   =    43
Gaps             =      2  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40        50        60
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTV
              || | |||||| | || ||| | || |||||||| |||||  |  || | ||||| | ||||
   LLLSPCLRGTPDCYFSHSPISSNFKVKFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTV
      110   X   120       130       140       150       160       170
```

```
        70        80        90       100       110       120       130
    AGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWI--TRQNFSRCLELQCQ
    |||||| ||| |||||||||  | ||| | ||||||||||  ||   |   ||  ||| |    ||||||||
    AGSKMQTLLEDVNTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLGLKPCIGKACQNFSRCLEVQCQ
      180       190       200       210       220       230       240

         140       150    X
    PDSSTLPPPWSPRPLEATAPTAP
    |||||| || ||  ||||    |
    PDSSTLLPPRSPIALEATELPEPRPRQLLLLLL
    250       260       270       280
```

2. US-08-162-407-6 (1-235)
   SSVHNF1 S.scrofa mRNA for vHNF-1.

```
LOCUS       SSVHNF1      2821 bp    RNA             MAM       17-APR-1993
DEFINITION  S.scrofa mRNA for vHNF-1.
ACCESSION   X69675
KEYWORDS    DNA binding protein; homeoprotein; nucleus; transcription factor;
            vHNF-1.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
            Theria; Eutheria; Artiodactyla; Suiformes; Suidae.
REFERENCE   1  (bases 1 to 2821)
  AUTHORS   Menoud,P.A., Matthies,R., Hofsteenge,J. and Nagamine,Y.
  TITLE     Purification and cDNA cloning of a transcription factor which
            functionally cooperates within a cAMP regulatory unit in the
            porcine uPA gene
  JOURNAL   Nucleic Acids Res. 21, 1845-1852 (1993)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 2821)
  AUTHORS   Nagamine,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1992) to the EMBL/GenBank/DDBJ databases. Y.
            Nagamine, Friedrich Miescher Institute, PO Box 2543, CH-4002 Basel,
            SWITZERLAND
  STANDARD  full automatic
COMMENT     NCBI gi: 2156
FEATURES             Location/Qualifiers
     source          1..2821
                     /organism="Sus scrofa"
                     /haplotype="2n"
                     /tissue_type="kidney"
                     /cell_type="epithelia"
                     /cell_line="LLC-PK-1"
                     /clone_lib="lambda-gt11 cDNA"
                     /clone="BluescriptKS"
     CDS             150..1829
                     /note="NCBI gi: 2157"
                     /codon_start=1
                     /product="porcine vHNF-1"
                     /translation="MVSKLTSLQQELLSALLSSGVTKEVLVQALEELLPSPSFGVKLE
                     TLPLSPGSGTEPDTKPVFHTLTNGHAKGRLSGDEGSEDGDDYDTPPILKELQALNTEE
                     AADDGAEVDRMLSEDPWRAAKMIKGYMQQHNIPQREVVDVTGLNQSHLSQHLNKGTPM
                     KTQKRAALYTWYVRKQREILRQFNQTVQSSGNLTDKSSQDQLLFLFPEFSQQSQGPGQ
                     SDDACSEPTNKKMRRNRFKWGPASQQILYQAYDRQKNPSKEEREALVEECNRAECLQR
                     GVSPSKAHGLGSNLVTEVRVYNWFANRRKEEEAFRQKLAMDAYSSNQTHSLNTLLSHS
```



```
                     SPHHQPSTSPPNKLPGVRYNQQGNNEVTSSSTISHHGNSAMVTSQSVLQQVSPASLDP
                     GHNLLSPDGKMQISVSGGGLPPVSTLTNIHSLSHHNPQQSQNLIMTPLSGVMAIAQSL
                     NSSQAQSVPVINSVAGSLAALQPVQFSQQLHSPHQQPLMQQSPGSHMAQQPFMAAVTQ
                     LQNSHMYTHKQEPPQYSHTSRFPPAMVVTDTSSISTLTNMSSSKQCPLQAW"
     misc_feature    150..245
                     /note="dimerization domain"
     misc_feature    718..890
                     /note="POU motif"
     misc_feature    891..1196
                     /note="homeodomain"
BASE COUNT      686 a    919 c    686 g    530 t
ORIGIN
```

```
Initial Score    =     21  Optimized Score  =     25  Significance =  9.52
Residue Identity =    25%  Matches          =     32  Mismatches   =    72
Gaps             =     22  Conservative Substitutions               =     0
Translation Frame=      1

       40    X   50        60            70        80        90       100
    SNLQDEELCGGLWRLVLAQRWMERLKTVAGSKMQG----LLERVNTEIHFVTKCAFQPPPSCLRFVQTNISR
                | |       |||   ||   |     |     |  |  |      |  | |       ||
              PGPRGV---TGVWRLKGGGGSWTWGSPVKLPSTLLTGTHPPT--PSPPSFSVL-GKWCPSSR
              X          10        20        30          40         50

      110       120       130       140                 150    X
    LLQETSEQLVALKPWITRQNFSRCLELQCQPDSSTLPPPWSPRPLEA----------TAPTAP
        |       |   |   ||     | |  |       | || |            |||
    RSNKNS-XAPCXAPGSPRRCWSRPWRSCCHPRVSGXSWRRCPCPLGAGLSRTPSRSSTLSPTATPRAACPGT
      60         70        80        90       100       110       120

    RA
```

3. US-08-162-407-6 (1-235)
   AIBDVVP2 Avian infectious bursal disease virus RNA for VP2

```
LOCUS       AIBDVVP2     1471 bp    RNA             VRL       11-MAY-1994
DEFINITION  Avian infectious bursal disease virus RNA for VP2 and (partial)VP4
            proteins.
ACCESSION   X54858
KEYWORDS    capsid protein.
SOURCE      infectious bursal disease virus.
  ORGANISM  infectious bursal disease virus
            Viridae; ds-RNA nonenveloped viruses; Birnaviridae.
REFERENCE   1  (bases 1 to 1471)
  AUTHORS   Vakharia,V.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1990) to the EMBL/GenBank/DDBJ databases.
            Vakharia V.N., Univ. of Maryland, Ctr. for Agricultural
            Biotechnology and VA-MD Regional College of Veterinary Medicine,
            Veterinary Science Drive, College Park, MD 20742, USA
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1471)
  AUTHORS   Vakharia,V.N., Ahamed,B. and He,J.
  TITLE     Use of polymerase chain reaction for efficient cloning of dsRNA
            segments of infectious bursal disease virus
  JOURNAL   Avian Dis. 36, 736-742 (1992)
  STANDARD  full automatic
COMMENT     Isolated from Chicken, subspecies Leghorn.
```

```
            NCBI gi: 486622
 FEATURES             Location/Qualifiers
      source          1..1471
                      /organism="Avian infectious bursal disease virus"
                      /strain="E/DEL"
                      /tissue_type="bursa"
                      /clone="pE.Del-2"
                      /map="large segment A"
      mRNA            1..1471
      CDS             52..>1471
                      /note="NCBI gi: 486623"
                      /codon_start=1
                      /product="VP2 structural protein and partial for VP4"
                      /translation="MTNLQDQTQQIVPFIRSLLMPTTGPASIPDDTLEKHTLRSETST
                      YNLTVGDTGSGLIVFFPGFPGSIVGAHYTLQSNGNYKFDQMLLTAQNLPASYNYCRLV
                      SRSLTVRSSTLPGGVYALNGTINAVTFQGSLSELTDVSYNGLMSATANINDKIGNVLV
                      GEGVTVLSLPTSYDLGYVRLGDPIPAIGLDPKMVATCDSSDRPRVYTITAADNYQFSS
                      QYQTGGVTITLFSANIDAITSLSVGGELVFKTSVQSLVLGATIYLIGFDGTAVITRAV
                      AANNGLTAGIDNLMPFNLVIPTNEITQPITSIKLEIVTSKSDGQAGEQMSWSASGSLA
                      VTIHGGNYPGALRPVTLVAYERVATGSVVTVAGVSNFELIPNPELAKNLVTEYGRFDP
                      GAMNYTKLILSERDRLGIKTVWPTREYTDFREYFMEVADLNSPLKIAGAFGFKDIIRA
                      IRRIAVPVVSTLFPPAAPVAHAI"
      mat_peptide     52..1407
                      /product="VP2 structural protein"
      mat_peptide     1408..>1471
                      /product="VP4 structural protein"
 BASE COUNT      403 a    390 c    365 g    313 t
 ORIGIN

Initial Score    =     19  Optimized Score  =     26  Significance =  8.40
Residue Identity =    18%  Matches          =     31  Mismatches   =   118
Gaps             =     18  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40                50
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGG-------LWRLVLAQR-
                     || |       ||        ||        ||          ||   |
    IDNLMPFNLVIPTNEITQPITSIKLEIVTSKSDGQAGEQMSWSASGSLAVTIHGGNYPGALRPVTLVAYERV
         310  X    320       330       340       350       360       370

        60        70        80        90       100       110       120
    WMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITR-QNFSR
          ||||    |         ||        |         |   |       |     | ||    |
    ATGSVVTVAGVSNFELIPNPELAKNLVTEYGRFDP------GAMNYTKLILSERDRLGIKTVWPTREYTDFR
       380       390       400              410       420       430       440

      130          140       150    X
    CLELQCQPDSSTL---PPPWSPRPLEATAPTAP
              | |         |    |
    EYFMEVADLNSPLKIAGAFGFKDIIRAIRRIAVPVVSTLFPPA
           450       460       470  X    480


4. US-08-162-407-6 (1-235)
   IBDLGVP2     Infectious bursal disease virus genomic RNA for VP

 LOCUS       IBDLGVP2     1506 bp ss-RNA             VRL       24-JUL-1992
 DEFINITION  Infectious bursal disease virus genomic RNA for VP2, partial
             sequence.
```



```
 ACCESSION   D10065 D01240
 KEYWORDS    host-protective immunogen; structural protein.
 SOURCE      Infectious bursal disease virus (strain Delaware variant E) cDNA to
             genomic RNA.
   ORGANISM  Infectious bursal disease virus
             Viridae; ds-RNA nonenveloped viruses; Birnaviridae.
 REFERENCE   1  (bases 1 to 1506)
   AUTHORS   Heine,H.-G., Haritou,M., Failla,P., Fahey,K. and Azad,A.
   TITLE     Sequence analysis and expression of the host-protective immunogene
             VP2 of a variant strain of infectious bursal disease virus which
             can circumvent vaccination with standard type 1 strains
   JOURNAL   J. Gen. Virol. 72, 1835-1843 (1991)
   STANDARD  full automatic
 COMMENT     Submitted (07-Nov-1991) to DDBJ by: Hans-Georg Heine
             CSIRO, Australian Animal Health Laboratory
             P.O.Bag 24
             Geelong VIC 3220
             Australia
             Phone:  052-26-5222
             Fax:    052-23-1424
             E mail: hans@aahl.dah.csiro.au.

             NCBI gi: 221889
 FEATURES             Location/Qualifiers
      source          1..1506
                      /organism="Infectious bursal disease virus"
      misc_feature    1..48
                      /note="5'PCR primer N527"
      CDS             19..>1506
                      /gene="vp2"
                      /note="NCBI gi: 221890"
                      /codon_start=1
                      /product="VP2"
                      /translation="MTNLSDQTQQIVPFIRSLLMPTTGPASIPDDTLEKHTLRSETST
                      YNLTVGDTGSGLIVFFPGFPGSIVGAHYTLQSNGNYKFDQMLLTAQNLPASYNYCRLV
                      SRSLTVRSSTLPGGVYALNGTINAVTFQGSLSELTDVSYNGLMSATANINDKIGNVLV
                      GEGVTVLSLPTSYDLGYVRLGDPIPAIGLDPKMVATCDSSDRPRVYTITAADNYQFSS
                      QYQTGGVTITLFSANIDAITSLSVGGELVFKTSVQSLVLGATIYLIGFDGTAVITRAV
                      AANNGLTAGIDNLMPFNLVIPTNEITQPITSIKLEIVTSKSDGQAGEQMSWSASGSLA
                      VTIHGGNYPGALRPVTLVAYERVATGSVVTVAGVSNFELIPNPELAKNLVTEYGRFDP
                      GAMNYTKLILSERDRLGIKTVWPTREYTDFREYFMEVADLNSPLKIAGAFGFKDIIRA
                      IRRIAVPVVSTLFPPAAPLAHAIGEGVDYLLGDEAQAASGTARAKL"
      misc_feature    1461..1500
                      /note="3'PCR primer N526"
      misc_feature    1501..1506
                      /note="HindIII site downstream of VP2 insert in plasmid
                      pIP201"
 BASE COUNT      405 a    402 c    379 g    320 t
 ORIGIN      large genomic segment A.

Initial Score    =     19  Optimized Score  =     26  Significance =  8.40
Residue Identity =    18%  Matches          =     31  Mismatches   =   118
Gaps             =     18  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40                50
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGG-------LWRLVLAQR-
                     || |       ||        ||        ||          ||   |
    IDNLMPFNLVIPTNEITQPITSIKLEIVTSKSDGQAGEQMSWSASGSLAVTIHGGNYPGALRPVTLVAYERV
         300 X     310       320       330       340       350       360
```

```
      60        70        80        90       100       110       120
   WMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITR-QNFSR
         ||||      |         ||       |         |   |       |     | ||     |
   ATGSVVTVAGVSNFELIPNPELAKNLVTEYGRFDP------GAMNYTKLILSERDRLGIKTVWPTREYTDFR
      370       380       390              400       410       420

     130          140       150    X
   CLELQCQPDSSTL---PPPWSPRPLEATAPTAP
             | |             |    |
   EYFMEVADLNSPLKIAGAFGFKDIIRAIRRIAVPVVSTLFPPA
 430       440       450       460 X     470
```

5. US-08-162-407-6 (1-235)
   SCINVE Yeast gene SUC2 encoding invertase.

```
LOCUS       SCINVE      2873 bp    DNA             PLN       12-SEP-1993
DEFINITION  Yeast gene SUC2 encoding invertase.
ACCESSION   V01311 K00540
KEYWORDS    invertase; signal peptide.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Plantae; Thallobionta; Eumycota; Hemiascomycetes;
            Endomycetales; Saccharomycetaceae.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Carlson,M., Taussig,R., Kustu,S. and Botstein,D.
  TITLE     The secreted form of invertase in Saccharomyces cerevisiae is
            synthesized from mRNA encoding a signal sequence
  JOURNAL   Mol. Cell. Biol. 3, 439-447 (1983)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 2389)
  AUTHORS   Taussig,R. and Carlson,M.
  TITLE     Nucleotide sequence of the yeast SUC2 gene for invertase
  JOURNAL   Nucleic Acids Res. 11, 1943-1954 (1983)
  STANDARD  full automatic
REFERENCE   3  (bases 1 to 484)
  AUTHORS   Sarokin,L. and Carlson,M.
  TITLE     Comparison of two yeast invertase genes: conservation of the
            upstream regulatory region
  JOURNAL   Nucleic Acids Res. 13, 6089-6103 (1985)
  STANDARD  full automatic
REFERENCE   4  (bases 1 to 2873)
  AUTHORS   Sarokin,L. and Carlson,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1985) to the EMBL/GenBank/DDBJ databases
  STANDARD  full automatic
COMMENT     NCBI gi: 3833
FEATURES             Location/Qualifiers
     source          1..2873
                     /organism="Saccharomyces cerevisiae"
     mRNA            745..>2581
                     /note="messenger RNA"
     CDS             785..2383
                     /note="invertase;  NCBI gi: 3834"
                     /codon_start=1
                     /translation="MLLQAFLFLLAGFAAKISASMTNETSDRPLVHFTPNKGWMNDPN
                     GLWYDEKDAKWHLYFQYNPNDTVWGTPLFWGHATSDDLTNWEDQPIAIAPKRNDSGAF
                     SGSMVVDYNNTSGFFNDTIDPRQRCVAIWTYNTPESEEQYISYSLDGGYTFTEYQKNP
                     VLAANSTQFRDPKVFWYEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKLESAFANEGFL
```



```
                     GYQYECPGLIEVPTEQDPSKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQ
                     SRVVDFGKDYYALQTFFNTDPTYGSALGIAWASNWEYSAFVPTNPWRSSMSLVRKFSL
                     NTEYQANPETELINLKAEPILNISNAGPWSRFATNTTLTKANSYNVDLSNSTGTLEFE
                     LVYAVNTTQTISKSVFADLSLWFKGLEDPEEYLRMGFEVSASSFFLDRGNSKVKFVKE
                     NPYFTNRMSVNNQPFKSENDLSYYKVYGLLDQNILELYFNDGDVVSTNTYFMTTGNAL
                     GSVNMTTGVDNLFYIDKFQVREVK"
     mutation        924
                     /note="mutant suc2-215 (amber)"
BASE COUNT      855 a    578 c    537 g    903 t
ORIGIN
```

```
Initial Score    =     19  Optimized Score  =     32  Significance =  8.40
Residue Identity =    23%  Matches          =     43  Mismatches   =    97
Gaps             =     46  Conservative Substitutions               =     0
Translation Frame=      1
```

```
            X         10        20        30        40        50
            QDCSFQ-HS-PISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERL-
                 | | ||   |    |       ||            ||    |    |    ||   ||
   WLHFYXIPKEPCFSCQLHSIQRSKGVLVXTFSKMDYDGCQITRLQNXNLLLXXLEVLEARICICQRRFLRLP
        420       430       440       450       460       470       480

          60        70        80        90       100       110       120
   ---------KTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITR-Q
             | |  | ||  |             ||   | || |      ||  |   |       |
   IRMSRFDXSPNXARS-FQILLGHV---------YFYQPRCTCWRFLQPIFCWILQWYSFXSVXQSIXSGRFW
        490       500                 510       520       530       540

          130       140                      150    X
   NFSRCL--ELQCQPDSSTLPPPWSPR----------------PLEATAPTAP
       ||   ||  |         | |                | |       |
   XGLLCLANFLQHXP-----NLRFSIRYCLGFKLGVQCLCPNXPMEIIHVFGPQVFFEHXISS
      550             560       570       580       590       600
```

6. US-08-162-407-6 (1-235)
   ATT13408 1474 Arabidopsis thaliana cDNA clone 39B11T7.

```
ID   ATT13408   standard; RNA; EST; 374 BP.
XX
AC   T13408;
XX
DT   06-MAR-1994 (Rel. 38, Created)
DT   06-MAR-1994 (Rel. 38, Last updated, Version 1)
XX
DE   1474 Arabidopsis thaliana cDNA clone 39B11T7.
XX
KW   EST.
XX
OS   Arabidopsis thaliana
OC   Eukaryota; Plantae; Embryobionta; Magnoliophyta; Magnoliopsida;
OC   Dilleniidae; Capparales; Brassicaceae.
XX
RN   [1]
RP   1-374
RA   Newman T.;
RT   ;
RL   Unpublished.
XX
CC   Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan
```

```
CC   State University MSU-DOE-PRL, Michigan State University,Plant
CC   Biology Bldg.,E. Lansing,Mi Tel: 517-353-0854 Fax: 517-353-9168
CC   Email: 22313tcn@ibm.cl.msu.edu. NCBI gi: 429777
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..374
FT                   /organism="Arabidopsis thaliana"
FT                   /note="Thale cress"
FT                   /clone="39B11T7"
FT                   /strain="var columbia"
XX
SQ   Sequence 374 BP; 100 A; 75 C; 82 G; 101 T; 16 other;

Initial Score    =     18  Optimized Score  =     26  Significance =  7.84
Residue Identity =    23%  Matches          =     34  Mismatches   =    80
Gaps             =     31  Conservative Substitutions               =     0
Translation Frame=      1

        X  10        20        30        40            50        60
    QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGGLWRL----VLAQRWMERLKTVAGSKMQ
         |   |              ||||       |    | |   | |       ||    |
         QRFRI-------------GLLQD----LQWNXXNREQCTLCWPLFXWHWHGQRQSXXLLISXIANGV
         X                   10            20        30        40        50

    70        80           90       100        110         120       130
    GLLERVNTEIHFVTKC---AFQPPPSCLRFVQTNISRL-LQETSEQLVAL--KPWITRQNFSRCLELQCQPD
     ||        |  ||   |       |  |      || |   |  | ||      |   |         |
    LLLTGRIKQEDFXLKCYKKAYKGIDPSLXXV----FRLELVSSSKALQALGXDRNTTLKLFGESCSFFAHPL
          . 60        70        80            90       100       110

       140      150
    SSTLPPPWSPRPLEAT
    | |
    SNTXXK
   120   X


7. US-08-162-407-6 (1-235)
   T04557       604 Arabidopsis thaliana cDNA clone 39B11T7.

 LOCUS       T04557        374 bp ss-mRNA            EST       14-APR-1993
 DEFINITION  604 Arabidopsis thaliana cDNA clone 39B11T7.
 ACCESSION   T04557
 KEYWORDS    EST.
 SOURCE      Thale cress clone=39B11T7 library=Lambda-PRL2 strain=var columbia
             vector=lambda Zip-Lox Rsite1=Sal Rsite2=Not Lambda PRL2 is a cDNA
             library derived from equal quantities of 4 pools of mRNA.  The mRNA
             sources were 1) 7 day germinated etiolated seedlings; 2) tissue
             culture grown roots; 3) staged plants half with 24 hour light
             cycle, half on 16 hr light, 8 hour dark- rosettes;  4) same plants
             as 3 but aerial tissue (stems, flowers and siliques.  The vector is
             BRL's lambda Zip-Lox.  The cDNA inserts were directionally cloned
             with Sal-Not arms using oligo dT primed cDNA.
   ORGANISM  Arabidopsis thaliana
             Eukaryota; Plantae; Embryobionta; Magnoliophyta; Magnoliopsida;
             Dilleniidae; Capparales; Brassicaceae.
 REFERENCE   1  (bases 1 to 374)
   AUTHORS   Newman,T.
   JOURNAL   Unpublished (1993)
```



```
   STANDARD  full automatic
 COMMENT
             Contact: Newman,T.
             MSU-DOE Plant Research Laboratory
             Michigan State University
             MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
             Lansing,Mi
             Tel: 517-353-0854
             Fax: 517-353-9168
             Email: 22313tcn@ibm.cl.msu.edu.

             NCBI gi: 315717
 FEATURES             Location/Qualifiers
      source          1..374
                      /organism="Arabidopsis thaliana"
                      /clone="39B11T7"
                      /strain="var columbia"
                      /note="Thale cress"
 BASE COUNT      100 a      75 c      82 g     101 t      16 others
 ORIGIN

Initial Score    =     18  Optimized Score  =     26  Significance =  7.84
Residue Identity =    23%  Matches          =     34  Mismatches   =    80
Gaps             =     31  Conservative Substitutions               =     0
Translation Frame=      1

        X  10        20        30        40            50        60
    QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGGLWRL----VLAQRWMERLKTVAGSKMQ
         |   |              ||||       |    | |   | |       ||    |
         QRFRI-------------GLLQD----LQWNXXNREQCTLCWPLFXWHWHGQRQSXXLLISXIANGV
         X                   10            20        30        40        50

    70        80           90       100        110         120       130
    GLLERVNTEIHFVTKC---AFQPPPSCLRFVQTNISRL-LQETSEQLVAL--KPWITRQNFSRCLELQCQPD
     ||        |  ||   |       |  |      || |   |  | ||      |   |         |
    LLLTGRIKQEDFXLKCYKKAYKGIDPSLXWV----FRLGLVSSSKHLQALGXDRNTTLKLFGESCSFFAHPV
           60        70        80            90       100       110

       140      150
    SSTLPPPWSPRPLEAT
    | |
    STTXCK
   120   X


8. US-08-162-407-6 (1-235)
   SFIURA5A     Sordaria macrospora orotate phosphoribosyl transfe

 LOCUS       SFIURA5A     1246 bp ds-DNA             PLN       16-AUG-1994
 DEFINITION  Sordaria macrospora orotate phosphoribosyl transferase (ura5) gene,
             complete cds.
 ACCESSION   M26957
 KEYWORDS    orotate phosphoribosyl transferase; ura5 gene.
 SOURCE      Sordaria macrospora (strain FGSC 4818) ATCC 60255 DNA.
   ORGANISM  Sordaria macrospora
             Eukaryota; Fungi; Ascomycota; Pyrenomycetes; Sordariales;
             Sordariaceae.
 REFERENCE   1  (bases 1 to 1246)
   AUTHORS   Le Chevanton,L. and Leblon,G.
   TITLE     The ura5 gene of the ascomycete Sordaria macrospora: Molecular
```

```
            cloning, characterization and expression in Escherichia coli
   JOURNAL  Gene 77, 39-49 (1989)
   STANDARD full automatic
 COMMENT    NCBI gi: 341555
 FEATURES             Location/Qualifiers
     source           1..1246
                      /organism="Sordaria macrospora"
                      /strain="FGSC 4818"
                      /sequenced_mol="DNA"
                      /tissue_type="ATCC 60255"
     CAAT_signal      199..203
                      /gene="ura5"
                      /note="putative"
     CDS              369..1067
                      /gene="ura5"
                      /note="NCBI gi: 530193"
                      /codon_start=1
                      /product="orotate phosphoribosyl transferase"
                      /translation="MAALRYKADFLKASIDGGVLKFGSFELKSKRISPYFFNAGDFYR
                      ADLLQAISTAYAKCIIEAHKSGQLDFDIVFGPAYKGIPLATAATDKLAQLDPETYGKI
                      CYSFDRKEAKDHGEGGNIVGAPLKGKRILIVDDVITAGTAKREAIAKIEKEGGIVAGI
                      VVALDRMEKLPAADGDDSKPGPSAMVSSARSTAIPIFAILTLDDIIEGMRGLASPEDV
                      KKTEEYRAKYKATD"
     polyA_signal     1086..1091
                      /gene="ura5"
 BASE COUNT      279 a    370 c    300 g    297 t
 ORIGIN

Initial Score    =     18  Optimized Score  =     31  Significance =  7.84
Residue Identity =    22%  Matches          =     39  Mismatches   =   109
Gaps             =     25  Conservative Substitutions               =     0
Translation Frame=      1

            X       10        20                30            40        50
            QDCSFQHSPISSDFAVKIREL-------SDYLLQDYPV-TVASN---LQDEELCGGLWRLVL
                  | |    ||      |       |   | | ||    ||            | |
    FGIFXPHIFFPRGPHRAHRP-MTGFAGLAPXLAHSXPTISLTQLKDFPVXRQESNPSAPPIFYFFLILSRFLP
      30      X 40          50        60        70        80        90

          60        70        80        90       100       110
    AQRWMERLKTVAGS-KMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKP------
    |     |       |     ||          ||       |||       |  |       |    |  |
    AFFQTSRFFLWDPSISFFKLLFFYRQWPLCVTRLISSRPPS------TAASXSSAASSSSPSAFPPTSSTPA
    100       110       120       130               140       150       160

     120       130       140       150    X
    WITRQNFSRCLELQCQPDSSTLPPPWSPRPLEATAPTAP
      |    ||   |     || |    |       ||
    TSTVPTSSRPSLLPTPSASSRLTRAASWTLTLSLAPPTRAFPSLPLPPI
        170       180       190       200 X     210


9. US-08-162-407-6 (1-235)
   RATNOSI      Rat nitric oxide synthase mRNA, complete cds.

 LOCUS       RATNOSI      3595 bp ss-mRNA            ROD       09-JUN-1993
 DEFINITION  Rat nitric oxide synthase mRNA, complete cds.
 ACCESSION   L12562
 KEYWORDS    nitric oxide synthase.
 SOURCE      Rattus norvegicus (strain Sprague-Dawley) adult cDNA to mRNA.
```



```
   ORGANISM  Rattus norvegicus
             Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
             Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
 REFERENCE   1  (bases 1 to 3595)
   AUTHORS   Wood,E.R., Berger,H.Jr.., Sherman,P.A. and Lapetina,E.G.
   TITLE     Hepatocytes and macrophages express an identical cytokine inducible
             nitric oxide synthase gene
   JOURNAL   Biochem. Biophys. Res. Commun. (1993) In press
   STANDARD  full automatic
 COMMENT     NCBI gi: 310187
 FEATURES             Location/Qualifiers
     source           1..3595
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /cell_line="primary culture"
                      /cell_type="hepatocyte"
                      /dev_stage="adult"
                      /sequenced_mol="cDNA to mRNA"
     CDS              103..3546
                      /note="inducible form; putative;  NCBI gi: 310188"
                      /codon_start=1
                      /product="nitric oxide synthase"
                      /translation="MACPWKFLFRVKSYQGDLKEEKDINNNVEKTPGAIPSPTTQDDP
                      KSHKHQNGFPQFLTGTAQNVPESLDKLHVTPSTRPQHVRIKNWGNGEIFHDTLHHKAT
                      SDISCKSKLCMGSIMNSKSLTRGPRDKPTPVEELLPQAIEFINQYYGSFKEAKIEEHL
                      ARLEAVTKEIETTGTYQLTLDELIFATKMAWRNAPRCIGRIQWSNLQVFDARSCSTAS
                      EMFQHICRHILYATNSGNIRSAITVFPQRTDGKHDFRIWNSQLIRYAGYQMPDGTIRG
                      DPATLEFTQLCIDLGWKPRYGRFDVLPLVLQAHGQDPEVFEIPPDLVLEVTMEHPKYE
                      WFQELGLKWYALPAVANMLLEVGGLEFPACPFNGWYMGTEIGVRDFCDTQRYNILEEV
                      GRRMGLETHTLASLWKDRAVTEINAAVLHSFQKQNVTIMDHHTASESFMKHMQNEYRA
                      RGGCPADWIWLVPPVSGSITPVFHQEMLNYVLSPFYYYQIEPWKTHIWQDEKLRPRRR
                      EIRFTVLVKAVFFASVLMRKVMASRVRATVLFATETGKSEALARDLAALFSYAFNTKV
                      VCMEQYKANTLEEEQLPLVVTSTFGNGDCPSNGQTLKKSLFMMKELGHTFRYAVFGLG
                      SSMYPQFCAFAHDIDQKLSHLGASQLAPTGEGDELSGQEDAFRSWAVQTFRAACETFD
                      VRSKHCIQIPKRYTSNATWEPEQYKLTQSPESLDLNKALSSIHAKNVFTMRLKSLQNL
                      QSEKSSRTTLLVQLTFEGSRGPSYLPGEHLGIFPGNQTALVQGILERVVDCSSPDQTV
                      CLEVLDESGSYWVKDKRLPPCSLRQALTYFLDITTPPTQLQLHKLARFATEETHRQRL
                      EALCQPSEYNDWKFSNNPTFLEVLEEFPSLRVPAAFLLSQLPILKPRYYSISSSQDHT
                      PSEVHLTVAVVTYRTRDGQGPLHHGVCSTWINNLKPEDPVPCFVRSVSGFQLPEDPSQ
                      PCILIGPGTGIAPFRSFWQQRLHDSQHRGLKGGRMTLVFGCRHRREDHLYQEEMQEMV
                      RKGVLFQVHTGYSRLPGKPKVYVQDILQKELADEVFSVLHGEQGHLYVCGDVRMARDV
                      ATTLKKLVAAKLNLSEEQVEDYFFQLKSQKRYHEDIFGAVFSYGAKKGNTLEEPKGTR
                      L"
 BASE COUNT      870 a   1044 c    956 g    725 t
 ORIGIN

Initial Score    =     18  Optimized Score  =     35  Significance =  7.84
Residue Identity =    24%  Matches          =     45  Mismatches   =   104
Gaps             =     35  Conservative Substitutions               =     0
Translation Frame=      1

            X       10        20            30        40        50
            QDCSFQHSPISSDFAVKIRELSDYLLQ---DYPVTVASNLQDEELCG--GLWRLVLAQRWME
                  |     | |         |  |   ||        |    |  |   |   |    |
   QYKLTQSPESLDLNKALSSIHAKNVFTMRLKSLQNLQSEKSSRTTLLVQLTFEGSRGPSYLPGEHLGIFPGN
 740       750       760       770       780       790       800       810

  60        70                80                90        100       110
 RLKTVAGSKMQGLLERV--------NTEIHFV---------TKCAFQPPPSCLRFVQT-NISRLLQETSEQLVA
     |     || ||||         |   |           |    || | ||   |         |  ||
```

```
QTALV-----QGILERVVDCSSPDQTVCLEVLDESGSYWVKDKRLPPCS-LRQALTYFLDITTPPTQLQLHK
          820       830       840       850       860       870

    120       130           140           150    X
LKPWITRQNFSRCLELQCQP----DSSTLPPPWSPRPLE----ATAPTAP
|   |     ||  |||    |     |     ||      | |
LARFATEETHRQRLEALCQPSEYNDWKFSNNPTFLEVLEEFPSLRVPAAFLLSQLPILKP
 880       890       900       910       920     X 930
```

10. US-08-162-407-6 (1-235)
    RATNWUT Rat mRNA.

```
LOCUS       RATNWUT      3614 bp ss-mRNA            ROD       14-JAN-1995
DEFINITION  Rat mRNA.
ACCESSION   D44591
KEYWORDS    .
SOURCE      Rattus norvegicus neutrophil cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
REFERENCE   1  (bases 1 to 3614)
  AUTHORS   Kosuga,K., Yui,Y., Hattori,R., Sase,K., Eizawa,H., Aoyama,T.,
            Inoue,R. and Sasayama,S.
  TITLE     Cloning of an inducible nitric oxide synthase from rat
            polymorphonuclear neutrophils
  JOURNAL   Endothelium 2, 217-221 (1994)
  STANDARD  full automatic
COMMENT     Submitted (29-Dec-1994) to DDBJ by:
            Kunihiko Kosuga
            Department of Internal Medicine(3rd Division)
            Kyoto University
            54 Kawaracho, Shogoin
            Sakyo-ku
            Kyoto 606-01
            Japan
            Tel:  075-751-3200
            Fax:  075-761-4616.

            NCBI gi: 624916
FEATURES             Location/Qualifiers
     source          1..3614
                     /organism="Rattus norvegicus"
                     /cell_type="neutrophil"
                     /sequenced_mol="cDNA to mRNA"
BASE COUNT      875 a   1056 c    950 g    733 t
ORIGIN
```

```
Initial Score    =     18  Optimized Score  =     35  Significance =  7.84
Residue Identity =    24%  Matches          =     45  Mismatches   =   104
Gaps             =     35  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20          30        40          50
             QDCSFQHSPISSDFAVKIRELSDYLLQ---DYPVTVASNLQDEELCG--GLWRLVLAQRWME
              |    | |       | |   ||       |    |  |   |   |    |
   QYKLTQSPEPLDLNKALSSIHAKNVFTMRLKSLQNLQSEKSSRTTLLVQLTFEGSRGPSYLPGEHLGIFPGN
        720 X     730       740       750       760       770       780

    60        70                80                90       100       110
```



```
RLKTVAGSKMQGLLERV-------NTEIHFV--------TKCAFQPPPSCLRFVQT-NISRLLQETSEQLVA
    |     || ||||       |    |         |    || | ||   |          | ||
QTALV-----QGILERVVDCSSPDQTVCLEVLDESGSYWVKDKRLPPCS-LRQALTYFLDITTPPTQLQLHK
         790       800       810       820       830       840

    120       130           140           150    X
LKPWITRQNFSRCLELQCQP----DSSTLPPPWSPRPLE----ATAPTAP
|   |     ||  |||    |     |     ||      | |
LARFATEETHRQRLEALCQPSEYNDWKFSNNPTFLEVLEEFPSLRVPAAFLLSQLPILKP
850       860       870       880       890       900
```

11. US-08-162-407-6 (1-235)
    MUSINOSL Mus musculus nitric oxide synthase (iNOS1) mRNA, c

```
LOCUS       MUSINOSL     3690 bp ss-mRNA            ROD       24-APR-1992
DEFINITION  Mus musculus nitric oxide synthase (iNOS1) mRNA, complete cds.
ACCESSION   M87039
KEYWORDS    cytosolic protein; nitric oxide synthase.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
REFERENCE   1  (bases 1 to 3690)
  AUTHORS   Xie,Q.-W., Cho,H.J., Calaycay,J., Mumford,R., Swiderek,K.M.,
            Lee,T.B., Ding,A., Troso,T. and Nathan,C.
  TITLE     Cloning and characterization of inducible nitric oxide synthase
            from mouse macrophage
  JOURNAL   Science 256, 225-228 (1992)
  STANDARD  full automatic
COMMENT     NCBI gi: 198406
FEATURES             Location/Qualifiers
     source          1..3690
                     /organism="Mus musculus"
                     /cell_line="RAW 264.7"
                     /cell_type="macrophage"
                     /sequenced_mol="cDNA to mRNA"
     5'UTR           1..255
                     /gene="iNOS1"
     CDS             256..3690
                     /gene="iNOS1"
                     /EC_number="1.14.23"
                     /note="NCBI gi: 198407"
                     /codon_start=1
                     /product="nitric oxide synthase"
                     /translation="MACPWKFLFKVKSYQSDLKEEKDINNNVKKTPCAVLSPTIQDDP
                     KSHQNGSPQLLTGTAQNVPESLDKLHVTSTRPQYVRIKNWGSGEILHDTLHHKATSDF
                     TCKSKSCLGSIMNPKSLTRGPRDKPTPLEELLPHAIEFINQYYGSFKEAKIEEHLARL
                     EAVTKEIETTGTYQLTLDELIFATKMAWRNAPRCIGRIQWSNLQVFDARNCSTAQEMF
                     QHICRHILYATNNGNIRSAITVFPQRSDGKHDFRLWNSQLIRYAGYQMPDGTIRGDAA
                     TLEFTQLCIDLGWKPRYGRFDVLPLVLQADGQDPEVFEIPPDLVLEVTMEHPKYEWFQ
                     ELGLKWYALPAVANMLLEVGGLEFPACPFNGWYMGTEIGVRDFCDTQRYNILEEVGRR
                     MGLETHTLASLWKDRAVTEINVAVLHSFQKQNVTIMDHHTASESFMKHMQNEYRARGG
                     CPADWIWLVPPVSGSITPVFHQEMLNYVLSPFYYYQIEPWKTHIWQNEKLRPRRREIR
                     FRVLVKVVFFASMLMRKVMASRVRATVLFATETGKSEALARDLATLFSYAFNTKVVCM
                     DQYKASTLEEEQLLLVVTSTFGNGDCPSNGQTLKKSLFMLRELNHTFRYAVFGLGSSM
                     YPQFCAFAHDIDQKLSHLGASQLAPTGEGDELSGQEDAFRSWAVQTFRAACETFDVRS
                     KHHIQIPKRFTSNATWEPQQYRLIQSPEPLDLNRALSSIHAKNVFTMRLKSQQNLQSE
                     KSSRTTLLVQLTFEGSRGPSYLPGEHLGIFPGNQTALVQGILERVVDCPTPHQTVCLE
                     VLDESGSYWVKDKRLPPCSLSQALTYFLDITTPPTQLQLHKLARFATDETDRQRLEAL
```

```
                    CQPSEYNDWKFSNNPTFLEVLEEFPSLHVPAAFLLSQLPILKPRYYSISSSQDHTPSE
                    VHLTVAVVTYRTRDGQGPLHHGVCSTWIRNLKPQDPVPCFVRSVSGFQLPEDPSQPCI
                    LIGPGTGIAPFRSFWQQRLHDSQHKGLKGGRMSLVFGCRHPEEDHLYQEEMQEMVRKR
                    VLFQVHTGYSRLPGKPKVYVQDILQKQLANEVLSVLHGEQGHLYICGDVRMARDVATT
                    LKKLVATKLNLSEEQVEDYFFQLKSQKRYHEDIFGAVFSYGAKKGSALEEPKATRL"
 BASE COUNT      893 a   1050 c    979 g    768 t
 ORIGIN

Initial Score    =     18  Optimized Score  =     34  Significance =  7.84
Residue Identity =    22%  Matches          =     42  Mismatches   =   108
Gaps             =     33  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40        50
             QDCSFQHSPISSDFAVKIRELSDYLLQ---DYPVTVASNLQDEELCG--GLWRLVLAQRWME
              |     | |       | |   ||       |    | |   |   |    |
   QYRLIQSPEPLDLNRALSSIHAKNVFTMRLKSQQNLQSEKSSRTTLLVQLTFEGSRGPSYLPGEHLGIFPGN
    790       800       810       820       830       840       850

    60        70          80                  90       100       110
   RLKTVAGSKMQGLLERV---NTEIHFV-----------TKCAFQPPPSCLRFVQTNISRLLQETSEQLVAL
       |     || ||||    |    |             |    || |               |  ||  |
   QTALV-----QGILERVVDCPTPHQTVCLEVLDESGSYWVKDKRLPPCSLSQALTYFLDITTPPTQLQLHKL
  860           870       880       890       900       910       920

      120       130           140           150    X
   KPWITRQNFSRCLELQCQP----DSSTLPPPWSPRPLE----ATAPTAP
       |       ||  |||     |       |     ||       | |
   ARFATDETDRQRLEALCQPSEYNDWKFSNNPTFLEVLEEFPSLHVPAAFLLSQLPILKP
     930       940       950       960       970   X   980


12. US-08-162-407-6 (1-235)
    RNU03699     Rattus norvegicus Sprague Dawley inducible nitric

 LOCUS       RNU03699     3943 bp    mRNA            ROD       06-AUG-1994
 DEFINITION  Rattus norvegicus Sprague Dawley inducible nitric oxide synthase
             (NOS) mRNA,  complete cds.
 ACCESSION   U03699
 KEYWORDS    .
 SOURCE      rat.
   ORGANISM  Rattus norvegicus
             Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
             Eutheria; Rodentia; Myomorpha; Muridae; Rattus.
 REFERENCE   1  (bases 1 to 3943)
   AUTHORS   Galea,E., Reis,D.J. and Feinstein,D.L.
   TITLE     Cloning and expression of inducible nitric oxide synthase from rat
             astrocytes
   JOURNAL   J. Neurosci. Res. 37, 406-414 (1994)
   STANDARD  full automatic
 REFERENCE   2  (bases 1 to 3943)
   AUTHORS   Feinstein,D.L.
   TITLE     Direct Submission
   JOURNAL   Submitted (23-NOV-1993) Douglas L. Feinstein, Neurobiology, Cornell
             University Medical College, 411 East 69th Street, New York, NY
             10021, USA
   STANDARD  full automatic
 COMMENT     NCBI gi: 430718
 FEATURES             Location/Qualifiers
      source          1..3943
```



```
                    /clone="pAstNOS1-4"
                    /clone_lib="Lambda Zap II-LPS treated astrocytes"
                    /strain="Sprague Dawley"
                    /organism="Rattus norvegicus"
                    /tissue_type="astrocytes"
      mRNA          1..3939
      CDS           4..3447
                    /standard_name="NOS"
                    /note="NCBI gi: 430719"
                    /codon_start=1
                    /product="nitric oxide synthase"
                    /translation="MACPWKFLFRVKSYQGDLKEEKDINNNVEKTPGAIPSPTTQDDP
                    KSHKHQNGFPQFLTGTAQNVPESLDKLHVTPSTRPQHVRIKNWGNGEIFHDTLHHKAT
                    SDISRKSKLCMGSIMNSKSLTRGPRDKHTPVEELLPQAIEFINQYYGSFKEAKIEEHL
                    ARLEAVTKEIETTGTYQLTLDELIFATKMAWRNAPRCIGRIQWSNLQVFDARSCSTAS
                    EMFQHICRHILYATNSGNIRSAITVFPQRTDGKHDFRIWNSQLIRIAGYQMPDGTIRG
                    EPATLEFTQLCIDLGWKPRYGRFDVLPLVLQAHGQDPEVFEIPPDLVLEVTMEHPKYE
                    WFQELGLKWYALPAVANMLLEVGGLEFPACPFNGWYMGTEIGVRDFCDTQRYNILEEV
                    GRRMGLETHTLASLWKDRAATEINAAVLHSFQKQNVTIMDHHTASESFMKHMQNEYRA
                    RGGCPADWIWLVPPVSGSITPVFHQEMLNYVLSPFYYYQIEPWKTHIWQDEKLRPRRR
                    EIRFTVLVKAVFFASVLMRKVMASRVRATVLFATETGKSEALARDLAALFSYAFNTKV
                    VCMEQYKANTLEEEQLLLVVTSTFGNGDCPSNGQTLKKSLFMMKELGHTFRYAVFGLG
                    SSMYPQFCAFAHDIDQKLSHLGASQLAPTGEGDELSGQEDAFRSWAVQTFRAACETFD
                    VRSKHCIQIPKRYTSNATWEPEQYKLTQSPESLDLNKALSSIHAKNVFTMRLKSLQNL
                    QSEKSSRTTLLVQLTFEGSRGPSYLPGEHLGIFPGNQTGLVQGILERVVDCSSPDQTV
                    CLEVLDESGSYWVKDKRLPPCSLRQALTYFLDITTPPTQLQLHKLARFATEETHRQRL
                    EALCQPSEYNDWKFSNNPTFLEVLEEFPSLRVPAAFLLSQLPILKPRYYSISSSLDHT
                    PSEVHLTVAVVTYRTRDGQGPLHHGVCSTWINNLKPEDPVPCFVRSVSGFQLPEDPSQ
                    PCILIGPGTGIAPFRSFWQQRLHDSQHRGLKGGRMTLVFGCRHPEEDHLYQEEMQEMV
                    RKGVLFQVHTGYSRLPGKPKVYVQDILQKELADEVFSVLHGEQGHLYVCGDVRMARDV
                    ATTLKKLVAAKLNLSEEQVEDYFFQLKSQKRYHEDIFGAVFSYGAKKGNTLEEPKGTR
                    L"
      3'UTR         3448..3939
      polyA_signal  one-of(3474..3479,3902..3907)
      misc_signal   one-of(3769,3792,3804,3878)
                    /function="cis-regulation of mRNA stability"
      polyA_site    3943
 BASE COUNT      949 a   1154 c   1001 g    839 t
 ORIGIN

Initial Score    =     18  Optimized Score  =     35  Significance =  7.84
Residue Identity =    24%  Matches          =     45  Mismatches   =   104
Gaps             =     35  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40        50
             QDCSFQHSPISSDFAVKIRELSDYLLQ---DYPVTVASNLQDEELCG--GLWRLVLAQRWME
              |     | |       | |   ||       |    | |   |   |    |
   QYKLTQSPESLDLNKALSSIHAKNVFTMRLKSLQNLQSEKSSRTTLLVQLTFEGSRGPSYLPGEHLGIFPGN
     710     X 720       730       740       750       760       770

    60        70          80                  90       100       110
   RLKTVAGSKMQGLLERV-------NTEIHFV--------TKCAFQPPPSCLRFVQT-NISRLLQETSEQLVA
       |     || ||||        |    |         |    || | ||   |         |  ||
   QTGLV-----QGILERVVDCSSPDQTVCLEVLDESGSYWVKDKRLPPCS-LRQALTYFLDITTPPTQLQLHK
   780           790       800       810       820       830       840

      120       130           140           150    X
   LKPWITRQNFSRCLELQCQP----DSSTLPPPWSPRPLE----ATAPTAP
   |    |       ||  |||    |      |     ||       | |
```

```
    LARFATEETHRQRLEALCQPSEYNDWKFSNNPTFLEVLEEFPSLRVPAAFLLSQLPILKP
       850       860       870       880       890  X    900


13. US-08-162-407-6 (1-235)
    MUSRGEETS    Mouse 18S ribosomal RNA gene 5' end, and 5' extern

 LOCUS       MUSRGEETS     4194 bp ds-DNA             ROD       15-JUN-1989
 DEFINITION  Mouse 18S ribosomal RNA gene 5' end, and 5' external transcribed
             spacer (5'ETS).
 ACCESSION   M20154
 KEYWORDS    18S ribosomal RNA; external transcribed spacer.
 SOURCE      Mouse DNA.
   ORGANISM  Mus musculus
             Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
             Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
 REFERENCE   1  (bases 1 to 4194)
   AUTHORS   Bourbon,H.-M.M., Michot,B., Hassouna,N., Feliu,J. and
             Bachellerie,J.-P.P.
   TITLE     Sequence and secondary structure of the 5' external transcribed
             spacer of mouse pre-rRNA
   JOURNAL   DNA 7, 181-191 (1988)
   STANDARD  full automatic
 COMMENT     NCBI gi: 200733
 FEATURES             Location/Qualifiers
      source          1..4194
                      /organism="Mus musculus"
      rRNA            4175..>4194
                      /note="18S ribosomal RNA"
 BASE COUNT      375 a   1373 c   1375 g   1071 t
 ORIGIN

Initial Score    =     18  Optimized Score  =     30  Significance =  7.84
Residue Identity =    24%  Matches          =     38  Mismatches   =    98
Gaps             =     22  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40        50        60
             QDCSFQHSPIS-SDFAVKIRELSDYLLQDYP-VTVASNLQDEELCGGLWRLVLAQRWMERLK
               |    |  | |       | |       |   |   |   |   |   |   |  ||
    RCIPFPSGVCTRCFLXASRCSWSVPGLSPRCLLLSWWWRSPFPGVPPVL---RLAVCLPVCVXEAR--ERGV
           390       400       410       420           430       440

           70        80        90       100       110       120       130
    TVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQ
             || ||          ||          |     ||        |  |    |         ||
    EERRRGKTPLLRRVRRPPRDXYACAXGWCXAVAAGVGKFLERLIAFPWGALRGLAFGGDR--------LQ--
    450       460       470       480       490       500

        140       150    X
    PDSSTLPPPWSPR-PLEATAPTAP
         || |   | |||   | |
    ----GLPCPRMLRMPLEENLPVAADPPARSPACW
      510       520       530


14. US-08-162-407-6 (1-235)
    T35584       EST87799 Homo sapiens cDNA 5' end similar to None.

 LOCUS       T35584         319 bp ss-mRNA             EST       04-JAN-1995
```



```
 DEFINITION  EST87799 Homo sapiens cDNA 5' end similar to None..
 ACCESSION   T35584
 KEYWORDS    EST.
 SOURCE      human library=Human Lung primer=M13 Reverse.
   ORGANISM  Homo sapiens
             Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 319)
   AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
             Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
             White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
             Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
             FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
             Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle Jr,P.S.,
             Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
             Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
             Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
             Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
             Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,
             Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
             Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
             Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
             Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
             Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
             Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
   TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
             Based Upon 52 Million Basepairs of cDNA Sequence
   JOURNAL   Unpublished (1995)
   STANDARD  full automatic
 COMMENT     Other_ESTs: THC9516
             Contact: Venter, JC
             The Institute for Genomic Research
             932 Clopper Rd, Gaithersburg, MD 20878
             Tel: 3018699056
             Fax: 3018699423
             Email: tdbinfo@tdb.tigr.org
             For clone availability, additional sequence and expression
             information related to this EST, please contact the TIGR Database
             (tdbinfo@tdb.tigr.org).

             NCBI gi: 617682
 FEATURES             Location/Qualifiers
      source          1..319
                      /organism="Homo sapiens"
                      /note="human"
 BASE COUNT       73 a     97 c     89 g     58 t      2 others
 ORIGIN

Initial Score    =     17  Optimized Score  =     23  Significance =  7.28
Residue Identity =    25%  Matches          =     28  Mismatches   =    74
Gaps             =      6  Conservative Substitutions               =     0
Translation Frame=      1

    X        10        20        30        40        50        60
    QDCSFQHS-PISSDFAVKIRELSDY-LLQDYPVTVASNL--QDEELCGGLWRLVLAQRWMERLKTVAGSKMQ
        |  | |     ||        |  |          ||    ||| | |  |||              | |
    GAAGFSSSGPAATSGAVLQAATGMYEQLKGEWNRKSPNLSKCGEEL-GRL-KLVLLELNFLPTTGTKLTKQQ
    X       10        20        30        40          50        60        70

    70        80        90       100   X   110
```

```
    GLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVAL
      | |   ||           ||  |        ||
    LILARNILEIGAQWSILRKDIPSFERYMAQLKCLLL
            80        90       100     X
```

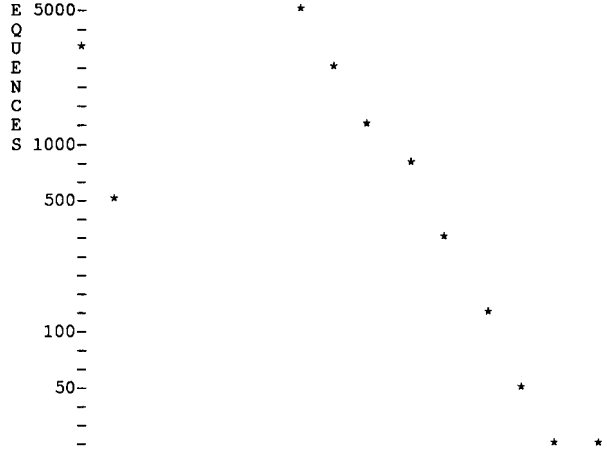
15. US-08-162-407-6 (1-235)
    CHKRERYF1   Chicken erythroid-specific transcription factor er

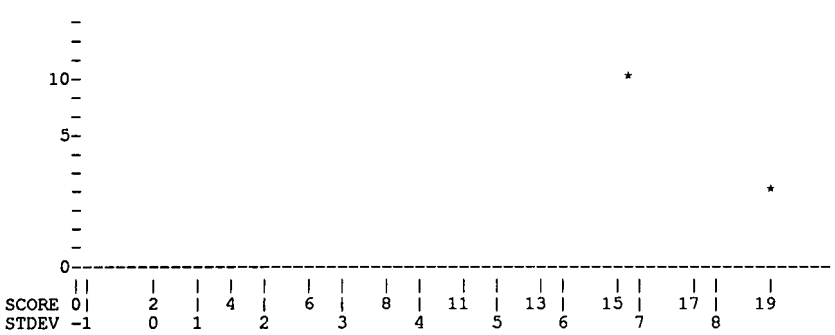
```
 LOCUS       CHKRERYF1    1068 bp ss-mRNA            VRT       15-DEC-1989
 DEFINITION  Chicken erythroid-specific transcription factor eryf1 mRNA,
             complete cds.
 ACCESSION   M26209
 KEYWORDS    DNA-binding transcription factor; globin.
 SOURCE      Chicken (White Leghorn) 9 day embryonic blood, cDNA to mRNA, clone
             E120-1.
   ORGANISM  Gallus gallus
             Eukaryota; Animalia; Chordata; Vertebrata; Aves; Neornithes;
             Neognathae; Galliformes; Phasianidae.
 REFERENCE   1  (bases 1 to 1068)
   AUTHORS   Evans,T. and Felsenfeld,G.
   TITLE     The erythroid-specific transcription factor Eryf1: A new finger
             protein
   JOURNAL   Cell 58, 877-885 (1989)
   STANDARD  full automatic
 COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
             by T.R.Evans, 17-JUL-1989.

             NCBI gi: 212628
 FEATURES             Location/Qualifiers
      source          1..1068
                      /organism="Gallus gallus"
      CDS             41..955
                      /note="Eryf1 protein;  NCBI gi: 212629"
                      /codon_start=1
                      /translation="MEFVALGGPDAGSPTPFPDEAGAFLGLGGGERTEAGGLLASYPP
                      SGRVSLVPWADTGTLGTPQWVPPATQMEPPHYLELLQPPRGSPPHPSSGPLLPLSSGP
                      PPCEARECVNCGATATPLWRRDGTGHYLCNACGLYHRLNGQNRPLIRPKKRLLVSKRA
                      GTVCSNCQTSTTTLWRRSPMGDPVCNACGLYYKLHQVNRPLTMRKDGIQTRNRKVSSK
                      GKKRRPPGGGNPSATAGGGAPMGGGGDPSMPPPPPPPPAAAPPQSDALYALGPVVLSGH
                      FLPFGNSGGFFGGGAGGYTAPPGLSPQI"
 BASE COUNT      206 a    371 c    342 g    149 t
 ORIGIN
```

```
Initial Score    =     17  Optimized Score  =     31  Significance =  7.28
Residue Identity =    24%  Matches          =     42  Mismatches   =    86
Gaps             =     45  Conservative Substitutions               =     0
Translation Frame=      1

             X        10          20        30        40        50
             QDCSFQHS-PISS---DFAVKIRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMER
               |   || |  |     |   ||        |     |         | |  |
    PPHGVRGAGGAGCGLPHSVPXXSRSLPGAGGGRE-------DGGGGAA---------GLLPPLRPRVPGAVG
           20 X      30        40                50                60

    60        70        80          90        100       110         120
    LKTVAGSKMQGLLERVNTEIHFVTKCA---FQPPPSCLRFVQTNISRLLQETSEQLVA--LKPWITRQNFS-
      |       |         |         |    ||||| ||          |   |   |     ||  |
    RHGYFGDPPVGAARHPNGAPPLPGAAATPPGQPPPSLLR----------APTATQQRAPTLRGPXVRQLRSD
     70        80        90        100                 110       120
```



```
              130       140       150    X
    --------RCLELQCQPDSSTLPPPWSPRPLEATAPTAP
            |   |  |     ||||  | |       ||
    GNAAVAPGRHRALPVQ-RLRALPPPQRPEPPAHPPQKAPAGEXARRHSV
  130       140       150       160       170
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq6nsqna.res made by  on Wed 5 Apr 95 18:27:19-PDT.


Query sequence being compared:US-08-162-407-6 (1-235)
Number of sequences searched:                57621
Number of scores above cutoff:                4182

     Results of the initial comparison of US-08-162-407-6 (1-235) with:
   Data bank : N-GeneSeq 17, all entries
```





PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Translation Frame	1		
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	1		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	3	3	2.00

Times:	CPU	Total Elapsed
	00:01:15.09	00:01:16.00

Number of residues: 8098565
Number of sequences searched: 57621
Number of scores above cutoff: 4182

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 8 standard deviations above mean ****					
1. Q14896	E/DEL-2 clone .	490	19	26	8.01	1
2. Q45441	Sequence of wild type Delawar	1060	19	26	8.01	1
	**** 6 standard deviations above mean ****					
3. Q26710	Gene regulating lipase expres	282	15	24	6.01	1
4. Q10571	Lipase modulating factor gene	344	15	27	6.01	1
5. Q05980	Sequence encoding Aleuria aur	456	15	28	6.01	1
6. Q44282	Lipase/lipase modulator fusio	706	15	27	6.01	1
7. Q06332	Murine IL-2R beta chain.	768	15	28	6.01	1
8. Q05873	Sequence encoding murine inte	768	15	29	6.01	1
9. Q41331	pMV206 - cassette contg. seq'	1373	15	31	6.01	1
10. Q57710	Neurocan DNA.	1730	15	30	6.01	1
	**** 5 standard deviations above mean ****					
11. N80626	Bovine Bone Morphogenic Prote	137	14	20	5.51	1
12. N92199	Region of clone bP-819 which	137	14	20	5.51	1
13. N81258	Encodes recombinant human pre	293	14	26	5.51	1
14. N71341	Sequence encoding rotavirus i	452	14	27	5.51	1
15. N90612	CD19 cDNA.	640	14	31	5.51	1
16. Q12071	T-PA with -ve charged finger	699	14	31	5.51	1
17. Q38723	AII/AVPv2 receptor coding seq	750	14	34	5.51	1
18. Q04339	Aquaricine I gene	757	14	17	5.51	1
19. Q53061	Human thyroid peroxidase.	1016	14	21	5.51	1
20. N90606	piH3M vector (ATCC 67,633) DN	1117	14	30	5.51	1
21. Q06313	Sequence of plasmid p267BGL.	1142	14	29	5.51	1
22. Q21166	Expression vector piH3M.	1300	14	30	5.51	1
23. Q06312	Sequence of plasmid p264BGL.	1326	14	29	5.51	1
24. Q05397	Secretion Vector pSEC-BP1 enc	1396	14	31	5.51	1
25. Q46823	pLISC-SE vector.	1505	14	31	5.51	1
26. Q51730	Plasmid pRK5.1 for huHGF site	1577	14	31	5.51	1
27. Q05590	Insert sequence encoding sign	1698	14	30	5.51	1
28. Q20732	pAD-CMV19 expression vector.	2020	14	30	5.51	1
29. Q20765	pAD-CMV19.	2020	14	30	5.51	1
30. Q20733	pAD-CMV1 expression vector.	2138	14	30	5.51	1
31. Q20766	pAD-CMV1.	2138	14	30	5.51	1
32. Q06283	pAD-CMV1 contg. a Tumour Necr	2138	14	30	5.51	1
33. N70996	Sequence encoding N-terminal	2749	14	30	5.51	1
34. Q55003	Humanised anti-CD18 Ab 60.3 v	3067	14	29	5.51	1
35. Q20907	DNA encoding the chimeric pro	3143	14	30	5.51	1
36. Q60158	Human brain Expressed Sequenc	88	13	22	5.01	1
37. N90627	Sequence encoding PB1rf HIV f	229	13	30	5.01	1
38. N81409	Beta chain gene of DP antigen	368	13	32	5.01	1
39. N80251	H-env probe for detection of	410	13	28	5.01	1
40. N70693	N-myc 1 proto-oncogene C-term	428	13	22	5.01	1
41. Q04459	N-myc 1 clone	428	13	22	5.01	1
42. N70450	Rotavirus SA-11 gene 6.	452	13	26	5.01	1
43. N90416	DNA sequence encoding methylt	626	13	27	5.01	1
44. Q47790	Bilirubin oxidase gene.	653	13	30	5.01	1
45. Q26657	glut4 promoter/enhancer .	773	13	28	5.01	1

1. US-08-162-407-6 (1-235)
   Q14896 E/DEL-2 clone .

ID Q14896 standard; DNA; 1471 BP.
AC Q14896;
DT 24-FEB-1992 (first entry)
DE E/DEL-2 clone .

```
KW   Infectious bursal disease virus; IBDV; vaccine; poultry; ss.
OS   Chicken infectious bursal disease virus.
FH   Key             Location/Qualifiers
FT   CDS             19..1470
FT   /*tag= a
FT   /note= "polypeptide encoded by E/DEL-2 clone"
PN   WO9116925-A.
PD   14-NOV-1991.
PF   30-APR-1991; U03056.
PR   04-MAY-1990; US-519202.
PA   (UYMA-) UNIV MARYLAND COL.
PI   Vakharia V.
DR   WPI; 91-353531/48.
DR   P-PSDB; R15053.
PT   Infectious bursal disease virus nucleic acids - useful as
PT   vaccines against IBDV in poultry
PS   Example; Page 28; 45pp; English.
CC   The sequence is that of a clone of variant Delaware type E of
CC   infectious bursal disease virus (IBDV). It can be used in broad
CC   spectrum IBDV poultry vaccines, administration is ophthalmically,
CC   by injection, nasally or orally, at any time after hatching and as a
CC   booster at other times. Poultry treated include chickens, roosters,
CC   broilers, roasters, breeders, layers, turkeys and ducks. See also
CC   Q14897.
SQ   Sequence   1471 BP;    403 A;    389 C;    366 G;    313 T;

Initial Score    =     19  Optimized Score  =     26  Significance =  8.01
Residue Identity =    20%  Matches          =     34  Mismatches   =   113
Gaps             =     22  Conservative Substitutions              =     0
Translation Frame=      1

             X       10        20        30        40              50
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGG-------LWRLVLAQR-
                     || |         ||          ||        ||         ||   |
   IDNLMPFNLVIPTNEITQPITSIKLEIVTSKSDGQAGEQMSWSASGSLAVTIHGGNYPGALRPVTLVAYERV
        310  X    320       330       340       350       360       370

      60        70        80        90       100       110       120
   WMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRL-LQETSEQLVALK-PWITR-QNF
        ||||     |          ||      |         |   | | |     |  |  | ||
   ATGSVVTVAGVSNFELIPNPELAKNLVTEYGRFDP------GAMNYTKLILSE--RDRVGIKTVWPTREYTD
      380       390       400              410       420         430

      130          140       150    X
   SRCLELQCQPDSSTL---PPPWSPRPLEATAPTAP
    |          | |             |    |
   FREYFMEVADLNSPLKIAGAFGFKDIIRAIRRIAVPVVSTLFPPA
  440       450       460       470  X    480


2. US-08-162-407-6 (1-235)
   Q45441       Sequence of wild type Delaware infectious bursal d

ID   Q45441 standard; DNA; 3180 BP.
AC   Q45441;
DT   07-OCT-1994  (first entry)
DE   Sequence of wild type Delaware infectious bursal disease virus
DE   (IBDV).
KW   Infectious bursal disease virus; IBDV; vaccine; monoclonal antibody
KW   Mab 67; assay; ss.
```



```
OS   Delaware infectious bursal disease virus.
FH   Key             Location/Qualifiers
FT   CDS             64..3102
FT   /*tag= a
PN   WO9406475-A.
PD   31-MAR-1994.
PF   01-SEP-1993; U08033.
PR   15-SEP-1992; US-944525.
PA   (UYMA-) UNIV MARYLAND BALTIMORE.
PI   Luetticken HD,  Snyder DB,  Vakharia V;
DR   WPI; 94-118171/14.
DR   P-PSDB; R51171.
PT   Attenuated live vaccine for Delaware strain infectious bursal
PT   disease virus - is produced by screening with a monoclonal
PT   antibody and used for immunising poultry against IBDV
PS   Claim 9; Fig 1A-1K; 24pp; English.
CC   The applicants have developed a Mab, designated Mab 67, which
CC   neutralises wild-type Delaware strain IBDV. It was developed
CC   according to the method set forth in US Patent 4,956,452. The
CC   hybridoma cell line expressing Mab 67 has been deposited under ATCC-
CC   HB111222. Bps 1-1404 of Q45441 is the nucleotide sequence of the
CC   gene responsible for the expression of Delaware IBDV structural
CC   protein(s) recognised by Mab 67. Mab 67 can be used in an assay to
CC   determine the presence of Delaware type IBDV and can be employed to
CC   prepare a vaccine conferring challenge protection against Delaware-
CC   type IBDV.
SQ   Sequence   3180 BP;    864 A;    880 C;    822 G;    614 T;

Initial Score    =     19  Optimized Score  =     26  Significance =  8.01
Residue Identity =    18%  Matches          =     31  Mismatches   =   118
Gaps             =     18  Conservative Substitutions              =     0
Translation Frame=      1

             X       10        20        30        40              50
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGG-------LWRLVLAQR-
                     || |         ||          ||        ||         ||   |
   IDNLMPFNLVIPTNEITQPITSIILEIVTSKSDGQAGEQMSWSASGSLAVTIHGGNYPGALRPVTLVAYERV
    310       320       330       340       350       360       370

      60        70        80        90       100       110       120
   WMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITR-QNFSR
        ||||     |          ||      |         |   |       |     | ||     |
   ATGSVVTVAGVSNFELIPNPELAKNLVTEYGRFDP------GAMNYTKLILSERDHLGIKTVWPTREYTDFR
  380       390       400       410              420       430       440

    130          140       150    X
   CLELQCQPDSSTL---PPPWSPRPLEATAPTAP
            | |             |    |
   EYFMEVADLNSPLKIAGAFGFKDIIRAIRRIAVPVVSTLFPPA
      450       460       470       480


3. US-08-162-407-6 (1-235)
   Q26710       Gene regulating lipase expression.

ID   Q26710 standard; DNA; 846 BP.
AC   Q26710;
DT   18-JAN-1993  (first entry)
DE   Gene regulating lipase expression.
KW   Fatty acid; fat; oil; ester; detergents; ss.
```

```
OS   Pseudomonas.
PN   J04169184-A.
PD   17-JUN-1992.
PF   31-OCT-1990; 294558.
PR   31-OCT-1990; JP-294558.
PA   (KURK ) KURITA WATER IND LTD.
DR   WPI; 92-253390/31.
DR   P-PSDB; R25719.
PT   Gene regulating expression of lipase - useful in prodn. of fatty
PT   acid from fats and fatty oils, prodn. of ester and addn. to
PT   detergents
PS   Claim 1; Fig 3; 9pp; Japanese.
CC   The gene regulating lipase expression was obtd. from a Pseudomonas
CC   clone, esp. from Pseudomonas KWI-56(FERM P-9659). Lipase is useful
CC   for the prodn. of fatty acid from fats and oils, the prodn. of
CC   esters, and for addition to biological detergents. Transformants
CC   contg. the lipase structural gene together with its regulatory gene
CC   can be used to produce 20-300 times greater lipase activity than
CC   transformants contg. the lipase structural gene alone.
SQ   Sequence   846 BP;    141 A;    296 C;    302 G;    107 T;

Initial Score    =     15  Optimized Score  =     24  Significance =  6.01
Residue Identity =    24%  Matches          =     32  Mismatches   =    77
Gaps             =     22  Conservative Substitutions               =     0
Translation Frame=      1

          30 X      40        50        60        70        80         90
    SDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTVAGSKMQGLLERVNTEIHFVTKC-AFQPPPS-
                ||        ||    ||    ||    ||    | |       |   |   |   |
               VPAS--------TGL-PPSLAGSSAPRLPLDAG----GHLAKSRAVRDFFDYCLTAQSDLSA
               X                 10        20              30        40

          100       110           120       130       140       150    X
    --CLRFVQTNISRLLQETSEQLVALKPW---ITRQNFSRCLELQCQPDSSTLPPPWSPRPLEATAPTAP
        ||    |   |   |   ||   |       |       | | |       |   |   |
    AGLDAFVMREIAAQLDGTVAQAEALDVWHRYRAYLDALAKLRDAGAADKSDL--GALQLALDQRASIAYRTL
   50        60        70        80        90        100         110     X

    GDWSQPF
  120
```

4. US-08-162-407-6 (1-235)
   Q10571 Lipase modulating factor gene, lim.

```
ID   Q10571 standard; DNA; 1032 BP.
AC   Q10571;
DT   15-APR-1991  (first entry)
DE   Lipase modulating factor gene, lim.
KW   Food; detergent; ss.
OS   Pseudomonas cepacia SB10.
PN   WO9100908-A.
PD   24-JAN-1991.
PF   24-JAN-1991; DK0170.
PR   07-JUL-1989; DK-003368.
PA   (NOVO ) NOVO NORDISK A/S.
PI   Jorgensen S;
DR   WPI; 91-051335/07.
DR   P-PSDB; R10864.
PT   DNA encoding lipase modulating factor acting in trans - used in
```



```
PT   prodn. of lipase derived from Pseudomonas cepacia, for increased
PT   prodn. of lipase.
PS   Claim 4; Fig 1; 57pp; English.
CC   The lim gene encodes a lipase modulating factor which acts in trans
CC   to modulate prodn. of P.cepacia lipase.  It is used to improve prodn.
CC   of recombinant lipase for use in e.g. the food industry or in
CC   detergents.    Chromosomal DNA from P. cepacia SB10 was used to
CC   prepare a library in phage lambda.   A positive phage, SJ150 was
CC   found to harbour plasmid pSJ150 contg. 6 kb of the lipase gene and
CC   the lipase modulating gene.  Sequencing showed that the codon TCG
CC   separates the lim start codon (ATG) from the lipase stop codon (TAA).
CC   See also Q10573.
SQ   Sequence   1032 BP;    150 A;    357 C;    401 G;    124 T;

Initial Score    =     15  Optimized Score  =     27  Significance =  6.01
Residue Identity =    20%  Matches          =     33  Mismatches   =   114
Gaps             =     15  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40        50        60
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTV
                                     | |       |   |   |             ||    ||
   VYGAVGLAAIAGVAMWSGAGRHGGTGASGEPPDASAARGP--AAAPPQAAVPASTSLPPSLAGSSAPRLPLD
      20     X  30        40        50          60        70        80

           70        80          90          100       110          120
   AGSKMQGLLERVNTEIHFVTKC-AFQPPPS---CLRFVQTNISRLLQETSEQLVALKPW---ITRQNFSRCL
   ||    | |         |   |   |   |       ||   |   |  |  |  || |             |
   AG----GHLAKARAVRDFFDYCLTAQSDLSAAGLDAFVMREIAAQLDGTVAQAEALDVWHRYRAYLDALAKL
          90        100       110       120       130       140       150

   130       140       150    X
   ELQCQPDSSTLPPPWSPRPLEATAPTAP
         | | |         |   |  |
   RDAGAVDKSDL--GALQLALDQRASIAYRWLGDWSQPF
       160         170       180
```

5. US-08-162-407-6 (1-235)
   Q05980 Sequence encoding Aleuria aurantia lectin.

```
ID   Q05980 standard; cDNA; 1370 BP.
AC   Q05980;
DT   14-JAN-1991  (first entry)
DE   Sequence encoding Aleuria aurantia lectin.
KW   Lectin; cancer; sugar chains; ss.
OS   Aleuria aurantia.
FH   Key             Location/Qualifiers
FT   CDS             80..1021
FT   /*tag=  a
FT   /label=A.aurantia lectin
PN   EP-387861-A.
PD   19-SEP-1990.
PF   14-MAR-1990; 104854.
PR   14-MAR-1989; JP-061280.
PA   (NICH-) NICHIREI CORP.
PI   Kobata A, Kochibe N, Fukumori F, Takeuchi N;
DR   WPI; 90-284100/38.
DR   P-PSDB; R06870.
PT   Cloned cDNA encoding Aleuria aurantia lectin - used for producing
```

```
PT   lectin for diagnosing cancer or sepg. or analysing sugar chains
PS   Disclosure; fig 1; 15pp; English.
CC   This cDNA encodes A.aurantia lectin which has a binding affinity
CC   for fucosyl linkages.  It recognises the change of oligosaccharide
CC   chains in cell surface glycoproteins and hence can be used as a
CC   diagnostic agent for e.g. cancer, where the oligosaccharide chains
CC   are modified.  It can also be used for sepg. or analysing various
CC   sugar chains.
SQ   Sequence   1370 BP;    312 A;    365 C;    373 G;    320 T;

Initial Score    =     15  Optimized Score  =     28  Significance =  6.01
Residue Identity =    22%  Matches          =     39  Mismatches   =    91
Gaps             =     47  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40        50        60
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTV
                 |     ||      |         | |   ||  |   || |    |          |  |
    LLAERXGGIVRGCLWRFMEHSSSDQGRI--------QAY-ALVARHLYCCELLGEHRHLCVLPSXRRLLAAV
       260    X  270       280               290       300       310

          70                    80        90       100           110
    AGSKMQGLLER-----------VN---TEIHFVTKCAFQPPPSCLRFVQTNISRL----LQETSEQLVALKP
    |    |||  |           ||      | |       | |  |      |||              | |
    AMDLRQGLVHRRGCSHWHSRGMVNGDQIRPHGV-------PTSVRRERLRHVSRLFYFCFLFSPYHAYAEK-
    320       330       340       350              360       370       380

     120          130       140       150    X
    WITRQNFSRCLEL----QCQPDSSTLPPPWSPRPLEATAPTAP
            | | |     |       |   |         |
    ------DGRKLSLFIWFICCVYGDFL--GWRSQGFFSVLGTWFILRLSLPWHT
                390       400         410     X 420
```

6. US-08-162-407-6 (1-235)
   Q44282 Lipase/lipase modulator fusion gene.

```
ID   Q44282 standard; DNA; 2118 BP.
AC   Q44282;
DT   09-DEC-1993  (first entry)
DE   Lipase/lipase modulator fusion gene.
KW   Lipase; LipA; LipD; lipase modulator; LimA; LimD; renaturation;
KW   chaperone molecule; lipolysis; detergent; denaturation; ss.
OS   Pseudomonas cepacia.
PN   WO9313200-A.
PD   08-JUL-1993.
PF   18-DEC-1992; DK0391.
PR   20-DEC-1991; WO-DK0402.
PA   (NOVO ) NOVO-NORDISK AS.
PI   Buckley CM,  Diderichsen BK,  Hobson A,  Joergensen ST;
PI   McConnell DJ;
DR   WPI; 93-227318/28.
PT   Prepn. of active lipase in high quantities - by subjecting to
PT   denaturation and restructuring in presence of chaperone molecule
PS   Claim 26; Page 47-48; 78pp; English.
CC   A DNA construct is claimed comprising a lipase coding sequence
CC   fused to a sequence coding for a chaperone molecule. One fusion
CC   construct is specifically claimed where the lip and lim coding
CC   sequences are both from Pseudomonas cepacia. The SEQ ID number for
CC   the specifically claimed fusion does not correspond to a nucleotide
```



```
CC   sequence so Q44282 (the only fusion gene sequence in the
CC   specification) is taken to be the intended fusion gene. Production
CC   of active lipase is enhanced by denaturation followed by
CC   renaturation in the presence of a lipase modulator chaperone
CC   protein.
SQ   Sequence   2118 BP;    336 A;    718 C;    772 G;    292 T;

Initial Score    =     15  Optimized Score  =     27  Significance =  6.01
Residue Identity =    20%  Matches          =     33  Mismatches   =   114
Gaps             =     15  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30        40        50        60
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTV
                                    |  |      |   |   |              ||     ||
    VYGAVGLAAIAGVAMWSGAGRHGGTGASGEPPDASAARGP--AAAPPQAAVPASTSLPPSLAGSSAPRLPLD
     380       390       400       410       420       430       440

          70        80        90       100       110       120
    AGSKMQGLLERVNTEIHFVTKC-AFQPPPS---CLRFVQTNISRLLQETSEQLVALKPW---ITRQNFSRCL
    ||    | |       |   |   |   |      ||   |   |  |  |  || |          |
    AG----GHLAKARAVRDFFDYCLTAQSDLSAAGLDAFVMREIAAQLDGTVAQAEALDVWHRYRAYLDALAKL
         450       460       470       480       490       500       510

    130       140       150    X
    ELQCQPDSSTLPPPWSPRPLEATAPTAP
          | | |       |   |  |
    RDAGAVDKSDL--GALQLALDQRASIAYRWLGDWSQPF
       520         530       540       550
```

7. US-08-162-407-6 (1-235)
   Q06332 Murine IL-2R beta chain.

```
ID   Q06332 standard; cDNA; 2306 BP.
AC   Q06332;
DT   07-FEB-1991  (first entry)
DE   Murine IL-2R beta chain.
KW   Interleukin; receptor; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             375..1994
FT   /*tag=  a
FT   misc_difference 1182..1184
FT   /*tag=  b
FT   /label=ggc, tgc
PN   AU9050726-A.
PD   13-SEP-1990.
PF   06-MAR-1990; 050726.
PR   07-MAR-1989; EP-104023.
PR   29-MAY-1989; EP-109656.
PR   20-JUL-1989; EP-113310.
PA   (BOEH ) BOEHRINGER INGELHEI.
PI   Taniguchi T, Hatakeyama M, Minamoto S, Kono T, Doi T;
PI   Miyasaka M, Tsudo, Karasuyma H;
DR   WPI; 90-327673/44.
DR   P-PSDB; R07507.
PT   Recombinant interleukin-2 receptor beta chain - useful for
PT   diagnosis and therapy by immune suppression or activation.
PS   Claim 4; Fig 8; 65pp; English.
```

```
CC   The sequence was obtd. from clones isolated from a cDNA library
CC   prepd. from RNA extracted from Concanavalin A-stimulated mouse
CC   spleen cells.  The clones were selected using a probe derived from
CC   the human sequence.  The sequence encodes a soluble portion of the
CC   IL-2R beta chain.
CC   See also Q06331.
SQ   Sequence   2306 BP;    487 A;    685 C;    580 G;    550 T;
SQ   4 Others;

Initial Score    =     15  Optimized Score  =     29  Significance =  6.01
Residue Identity =    20%  Matches          =     35  Mismatches   =   114
Gaps             =     24  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20                  30        40
             QDCSFQHSPISSDFAVKIRE----------LSDYLL---QDYPVTVASNLQDEELCGG----L
               |       |    |  |          |   |   |  |    |       | |
   ASSRDRRGSSWRCXSLVPHMRSRXGSKLNETIPGPGVPGASPXPLGQGQQIPXRRSS------PCHGSDTFC
          320       330       340       350       360               370

     50          60        70        80        90       100       110
   WRLV--LAQRWMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKP
   | ||  ||          |   |                  ||          ||     |   |||
   WSLVVFLASSPASTFWSRHGTLGHGXRQFSSATSQILLSSSPSXAPSMGETFRNGSPRLSPCPSSAPVALPL
    380       390       400       410       420       430       440

    120       130       140       150    X
   WITRQNFSRCLELQCQPDSSTLPPPWSPRPLEATAPTAP
     |   |       |         |  |||     | |
   RSLRWKCSTEIPRPCSCSCYRRTLPLYPRPAATHRPAASPTRATSSSIC
  450       460       470       480       490
```

8. US-08-162-407-6 (1-235)
   Q05873 Sequence encoding murine interleukin-2 receptor be

```
ID   Q05873 standard; DNA; 2306 BP.
AC   Q05873;
DT   07-JAN-1991  (first entry)
DE   Sequence encoding murine interleukin-2 receptor beta-chain.
KW   IL-2R beta; pIL-2R beta 9; ss;
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             375..1994
FT   /*tag=  a
PN   EP-386304-A.
PD   12-SEP-1990.
PF   29-MAY-1989; 109656.
PR   07-MAR-1989; EP-104023.
PR   29-MAY-1989; EP-109656.
PA   (OSAU ) OSAKA UNIVERSITY.
PI   Taniguchi T;
DR   WPI; 90-276456/37.
DR   P-PSDB; R06647.
PT   Recombinant interleukin-2 receptor beta chain - used for studying
PT   IL-2 system and producing antibodies for diagnosis and therapy.
PS   Claim 6; Fig 8; 37pp; English.
CC   IL-2 receptor beta chains are useful in studying the biochemistry of
CC   interleukin, and in diagnosis and therapy by immune suppression and
CC   activation.
```



```
SQ   Sequence   2306 BP;    482 A;    688 C;    585 G;    551 T;

Initial Score    =     15  Optimized Score  =     29  Significance =  6.01
Residue Identity =    19%  Matches          =     34  Mismatches   =   117
Gaps             =     20  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20        30                  40
             QDCSFQHSPISSDFAVKIRELSDYLLQDYPVTVAS----------NLQDEELCGG----LWR
               |       |    |  |          |  ||                | |     |
   ASSRDRRGSSWRCXSLVPHMRSRXGSKLNE----TIPGPGVPGASPXPFGQGQQIPXRRSSPCHGSDTFCWS
          320       330       340           350       360       370

     50          60        70        80        90       100       110
   LV--LAQRWMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWI
   ||  ||          ||    |                    ||          ||     |   |||
   LVVFLASSPASTFWSSAGTLGHGXRQFSSATSQILLSSSPSXAPSMGETFRNGSPRLSPCPSSAPVALPLRS
  380       390       400       410       420       430       440       450

    120       130       140       150    X
   TRQNFSRCLELQCQPDSSTLPPPWSPRPLEATAPTAP
    |   |       |         |  |||     | |
   LRWKCSTEIPRPCSCSCYRRTLPLYPRPAATHRPAASPTRATSSSIC
          460       470       480       490
```

9. US-08-162-407-6 (1-235)
   Q41331 pMV206 - cassette contg. seq's necessary for repli

```
ID   Q41331 standard; DNA; 4119 BP.
AC   Q41331;
DT   04-JUN-1993  (first entry)
DE   pMV206 - cassette contg. seq's necessary for replicn. in mycobacteria.
KW   cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;
KW   Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;
KW   pertussis; malaria; influenza virus; CTL; herpes virus.
OS   Mycobacterium.
FH   Key             Location/Qualifiers
FT   misc_feature    122..125
FT   /*tag=  a
FT   /codon= kan start codon
FT   misc_feature    935..937
FT   /*tag=  b
FT   /codon= kan stop codon.
PN   WO9307897-A.
PD   29-APR-1993.
PF   21-OCT-1992; U09075.
PR   21-OCT-1991; US-780261.
PA   (MEDI-) MEDIMMUNE INC.
PI   Stover CK;
DR   WPI; 93-152187/18.
PT   Expression vector for expressing protein or polypeptide in
PT   mycobacterium - contg DNA sequences encoding lipoprotein
PT   secretion signal and peptide heterologous to bacteria expressing
PT   fusion protein of lipoprotein heterologous to bacteria
PS   Example 1; Fig 12; 86pp; English.
CC   This sequence represents that of plasmid pMV206. This is a plasmid
CC   containing sequences necessary for replication in mycobacteria.
CC   It contains a synthetic multiple cloning site to facilitate
CC   versatile molecular cloning and manipulations for foreign gene
```

```
CC   expressions in mycobacteria and for integration into the
CC   mycobacterial chromosome.
SQ   Sequence   4119 BP;    922 A;    1129 C;    1163 G;    904 T;
SQ   1 Others;

Initial Score    =     15  Optimized Score  =     31  Significance =  6.01
Residue Identity =    22%  Matches          =     44  Mismatches   =    91
Gaps             =     58  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20                  30        40        50
             QDCSFQHSPISSDFAVKIREL----------SDYL--LQDYPVTVASNLQDEELCGGLWRLV
                    |    |         |          ||  |   ||  ||      | |   |
   PDQELPTLFPKVTGFSRAQIPNTVLLVXPXLGHHFKNSVAPPTYLALLILLPVAAAS-GDKSCLTGLDSRRX
 450       460       470       480       490       500       510

            60        70         80        90         100
   L---AQRWMERLKTVAGSKMQGLLER-VNTEIHFVTKCAFQPPPSCLRFVQ--TNISRLLQE----------
   |   ||    |    |||  |  |||   ||               ||  |     |  | |
   LPDKAQ----RSGXTAGSCTQPSLERTTYTE--------------LRYLQREHXESATLPEGRKADRYPVS
 520           530       540                     550       560       570

       110       120             130       140       150    X
   ----TSEQLVALKPWITRQNFS-----RCL--ELQCQPDSSTLPPPWSPRPLEATAPTAP
       | |   |          |    | |  |      ||      |       |     |
   GRVRTGERTRELPGGNAWYLYSPVGFRHLXLERRFLXCSSGGRSLWKNASNAAFLRSWPFAGLLLTCSF
       580       590       600       610       620       630       640


10. US-08-162-407-6 (1-235)
    Q57710       Neurocan DNA.

ID   Q57710 standard; DNA; 5191 BP.
AC   Q57710;
DT   11-AUG-1994  (first entry)
DE   Neurocan DNA.
KW   Neurocan; cell adhesion; leukocyte-endothelial cell recognition;
KW   tissue-related inflammation allergy; cellular; humoural;
KW   hypersensitivity; trauma; neuronal development; cell transport;
KW   infection; diagnosis; lectin; versican; aggrecan; gelsolin;
KW   receptor; cell recognition; membrane cytoplasmic protein; lipid;
KW   carbohydrate; saccharide; nucleoside; enzyme; ion; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             77..3850
FT   /*tag=  a
FT   /product= Neurocan
PN   WO9403601-A.
PD   17-FEB-1994.
PF   03-AUG-1993; U07306.
PR   03-AUG-1992; US-922911.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Margolis RK,  Margolis RU,  Rauch U;
DR   WPI; 94-065690/08.
DR   P-PSDB; R46627.
PT   Eukaryotic neurocan polypeptide(s) with epidermal growth factor,
PT   lectin or complement binding activity - used in the diagnosis,
PT   treatment or research of hypersensitivity and allergic diseases
PS   Claim 1; Page 69-75; 105pp; English.
CC   This sequence encodes a neurocan polypeptide.  This protein has
```



```
CC   several biological activities, including cell adhesion, leukocyte-
CC   endothelial cell recognition, tissue-related inflammation allergies,
CC   cellular and/or humoural hypersensitivity, trauma, neuronal
CC   development, and cell transport and/or infection.  Compositions
CC   containing them can be used as modulators of these conditions, and
CC   may be used as therapeutic, diagnostic, and/or research tools.
CC   Neurocan peptides can be used to mimic proteins, such as lectins,
CC   cell adhesion molecules, versicans, aggrecans or gelsolins, as
CC   receptor or effector subtypes.  The protein can be used to treat
CC   diseases involving a qualitative or quantitative pathological
CC   abnormality of cell adhesion or leukocyte-endothelial cell recognition,
CC   or a functionally associated molecule such as a membrane cytoplasmic
CC   protein, lipid, carbohydrate, saccharide, nucleoside, enzyme or ion.
SQ   Sequence   5191 BP;    1104 A;    1534 C;    1457 G;    1096 T;

Initial Score    =     15  Optimized Score  =     30  Significance =  6.01
Residue Identity =    23%  Matches          =     40  Mismatches   =    90
Gaps             =     40  Conservative Substitutions               =     0
Translation Frame=      1

             X       10        20            30        40            50
             QDCSFQHSPISSDFAVKIRELSDYLLQ---DYPVTVASNLQDEEL---CGGLWRLVLAQRWM
                    |   | ||         ||    |    |    |   |   ||          |
   LXLQPPLRLQEGYSAVWAPSSS-----GECLSCWCAQGQVQCPCHCAIPVXXRILPAPCG--------YHPM
    1170      1180           1190      1200      1210      1220

     60        70        80        90        100       110       120
   ERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQET-SEQLVALKPWITRQNFSRCL
       |         |         ||   |  |||       |       ||| |  |         |
   PKQWEV------GPASDCVHQAQAVTSDASTPPPSTPASQATQGAQKTQETPSGRL----GERXRGFLLTIQ
     1230           1240      1250      1260      1270           1280

   130       140             150    X
   ELQCQPDSSTLPPPWSPRPLE--------ATAPTAP
     | |    | ||  ||   |          | ||
   TNQAQ--APTPPPEHSPGDPEPRQPPERGWGHPAAPHPCSRSSVQS
   1290        1300      1310      1320      1330


11. US-08-162-407-6 (1-235)
    N80626       Bovine Bone Morphogenic Protein-3 in bp-819 detect

ID   N80626 standard; DNA; 413 BP.
AC   N80626;
DT   08-OCT-1990  (first entry)
DE   Bovine Bone Morphogenic Protein-3 in bp-819 detected by probe #2.
KW   Bone morphogenic protein; bBMP-3; probes; cartilage formation;
KW   bone formation; osteogenic cpds.;  prodontal disease; ss.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   CDS             41..374
FT   /*tag=  a
FT   /product=bBMP-3
FT   misc_RNA        274..298
FT   /*tag=  b
FT   /product=tryptic fragment 12
FT   misc_RNA        322..352
FT   /*tag=  C
FT   /product=tryptic fragment 10
PN   WO8800205-A.
```

```
PD   14-JAN-1988.
PF   30-JUN-1987; U01537.
PR   26-MAR-1987; US-031346,
PA   (GENE-) Genetics Inst Inc.
PI   Wozney JM, Rosen VA;
DR   WPI; 88-021565/03.
DR   P-PSDB; P81515.
PT   Bone morphogenic proteins - obtd. using recombinant DNA and used
PT   for inducing cartilage and bone formation.
PS   Disclosure; 7pp; English.
CC   This region of bp-819 hybridises to probe #2. The Bp-819 clone
CC   encodes the bovine bone growth factor bBMP-3.
CC   BBMP-3 is characterised by the sequences of clone bp-819  of N80626
CC   and N80627. The peptide is 175 amino acids long.
CC   This sequence is also used to screen a human genomic library to
CC   detect  the BMP-3 human bone inductive factor which is significantly
CC   homologous to the bovine factor.
CC   See also N80619-N80636 and N81963-64.
SQ   Sequence   413 BP;   119 A;   95 C;   111 G;   88 T;

Initial Score    =     14  Optimized Score  =     20  Significance =  5.51
Residue Identity =    22%  Matches          =     25  Mismatches   =    61
Gaps             =     23  Conservative Substitutions              =     0
Translation Frame=      1

             X       10              20          30              40
             QDCSFQHSPISSDFA------VKIRE-LSDYLLQDYPV--------TVASNLQDEELCGGLW
             ||        ||| |       |  |         | ||               |||
   GXSMGGEEALQD--------SSDSAPXXEXEQKETEEGTSAEESDAPVXXTDPEEGKKKAMDXTPELCQTVP
     30      X         40         50         60         70         80         90

    50         60         70         80         90   X   100
   RLVLAQRWMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLL
      |   | ||           || |      |       |   ||
   XSGLRRYWLERMDYFPQVLRCLLLLRSVPVPHAKGSHCFLSCPSHFH
          100        110        120        130      X
```

12. US-08-162-407-6 (1-235)
    N92199 Region of clone bP-819 which hybridzes with bovine

```
ID   N92199 standard; cDNA; 413 BP.
AC   N92199;
DT   06-APR-1990  (first entry)
DE   Region of clone bP-819 which hybridzes with bovine BMP-3 probe.
KW   BMP-3 protein; lambda bP-819; bone formation; tissue repair;
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             41..373
FT   /*tag= a
PN   WO8910409-A.
PD   02-NOV-1989.
PF   07-APR-1989; U01464.
PR   08-APR-1988; US-179197.
PA   (GENE-) Genentic Institute, Inc.
PI   Wang EA, Wozney JM, Rosen V;
DR   WPI; 89-339971/46.
DR   P-PSDB; P93233.
PT   Isolated DNA sequence and purified BMP-3 protein prepn. - by culturing
PT   cell transformed with defined DNA and recovery of 96 amino acid protein.
```



```
PS   Disclosure; page 20; 37pp; English.
CC   The sequence encodes part of a BMP-3 protein . The CDS is joined to the
CC   CDS of N92200. The sequence was isolated from clone lambda bP-819 by
CC   hybridization to a probe (N93420). Purified BMP-3 may be used to induce
CC   bone and/or cartilage formation and to promote wound healing and tissue
CC   repair.
SQ   Sequence   413 BP;   119 A;   95 C;   111 G;   88 T;

Initial Score    =     14  Optimized Score  =     20  Significance =  5.51
Residue Identity =    22%  Matches          =     25  Mismatches   =    61
Gaps             =     23  Conservative Substitutions              =     0
Translation Frame=      1

             X       10              20          30              40
             QDCSFQHSPISSDFA------VKIRE-LSDYLLQDYPV--------TVASNLQDEELCGGLW
             ||        ||| |       |  |         | ||               |||
   GXSMGGEEALQD--------SSDSAPXXEXEQKETEEGTSAEESDAPVXXTDPEEGKKKAMDXTPELCQTVP
     30      X         40         50         60         70         80         90

    50         60         70         80         90   X   100
   RLVLAQRWMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLL
      |   | ||           || |      |       |   ||
   XSGLRRYWLERMDYFPQVLRCLLLLRSVPVPHAKGSHCFLSCPSHFH
          100        110        120        130      X
```

13. US-08-162-407-6 (1-235)
    N81258 Encodes recombinant human preproapoprotein A1

```
ID   N81258 standard; cDNA; 881 BP.
AC   N81258;
DT   24-OCT-1990  (first entry)
DE   Encodes recombinant human preproapoprotein A1
KW   human preproapoprotein A1; high density lipoprotein deficiency; ss.
OS   synthetic.
FH   Key             Location/Qualifiers
FT   misc_RNA        20..73
FT   /*tag= a
FT   /note="encodes precursor peptide"
FT   misc_RNA        74..94
FT   /*tag= b
FT   /note="encodes propeptide"
FT   misc_RNA        95..823
FT   /*tag= c
FT   /note="encodes mature apoprotein A1"
FT   misc_difference 74..136
FT   /*tag= d
FT   /note="replaces wild type sequence to eliminate hairpins"
PN   EP-293357-A.
PD   30-NOV-1988.
PF   24-MAY-1988; 870095.
PR   28-MAY-1987; GB-012540.
PA   (UNIO) UCB SA.
PI   Bollen A, Gobert J, Wulfert E;
DR   WPI; 88-339891/48.
DR   P-PSDB; P80668.
PT   New DNA encoding human preproapoprotein A1 -
PT   modified to eliminate hairpin structures
PS   Claim 1; page 12; 25pp; French.
CC   The DNA fragment used to replace the wild-type sequence encoding
```

```
CC   amino acids -6 to +14 (= nucleotides 74 to 136) includes an
CC   additional ATG start codon and modifies codons corresponding to
CC   wild-type amino acids -6, -1, 1, 3, 4, 5, 6, 7, 10, 11 and 14.
CC   The changed codons still encode the same amino acids as in the
CC   wild-type protein but reduce formation of secondary structures in
CC   mRNA.
CC   See also N82064 for wild-type sequence.
SQ   Sequence   881 BP;   200 A;   269 C;   280 G;   132 T;

Initial Score    =     14  Optimized Score  =     26  Significance =  5.51
Residue Identity =    23%  Matches          =     34  Mismatches   =    79
Gaps             =     29  Conservative Substitutions               =     0
Translation Frame=      1

          20 X      30        40         50         60         70         80
    SDFAVKIRELSDYLLQDYPVTVASNLQDEELCG-GLWRLVLAQRWMERLKTVAGSKMQGLLERVNTEIHFVT
               |   |||           | ||  | |             ||       |       | |
               SPTALQD------------ESCGADLGRALPDGEPGYETFLAAGRTSTISLGXSXGLGHCVR
               X                    10        20         30        40        50

         90       100       110       120       130              140
    KCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQ------CQPDSSTL-PPPWSPRPL
      || |     |  |    ||             ||    |   | | ||          |   |  |    |
    GCA-QRQRQRL-CVPVXRLRL-------GKTAKPKAPXQLGQRDLHLQQAARTARPCDPGVLGXPGKGDRGP
            60                  70        80        90        100       110

    150    X
    EATAPTAP
    ||
    EAGDEQGSGGGEGQGAAL
        120       130
```

14. US-08-162-407-6 (1-235)
    N71341 Sequence encoding rotavirus inner shell protein VP

```
ID   N71341 standard; DNA; 1357 BP.
AC   N71341;
DT   24-MAY-1991  (first entry)
DE   Sequence encoding rotavirus inner shell protein VP6 serotype SAII.
KW   Diarrhoea; gastrointestinal disorder; RNA virus; vaccine; ss.
OS   Rotavirus.
FH   Key             Location/Qualifiers
FT   CDS             24..1217
FT   /*tag= a
PN   AU8666987-A.
PD   02-JUL-1987.
PF   23-DEC-1986; 017981.
PR   26-DEC-1985; US-813661.
PR   03-SEP-1986; US-903325.
PA   (UYSA-) UNIV OF SASKATCHEWA.
PI   Sabara MIJ, Frenchick PJ, Potter AA, Ijaz MK, Gilchrist JE;
DR   WPI; 87-228567/33.
DR   P-PSDB; P71563.
PT   New peptide fragments of rota:viral proteins - useful conjugates
PT   in vaccines for protecting against gastrointestinal disorders and
PT   diarrhoea
PS   Disclosure; Fig 2; 84pp; English.
CC   The peptide fragments of glycoprotein VP7 and proteins VP6 and VP3
CC   of rotaviruses are useful when attached to carriers as vaccines for
```



```
CC   birds and mammals, including man. The vaccines confer protection
CC   against gastrointestinal disorders and diarrhoea produced by the
CC   rotaviruses. For use in vaccines the peptides are covalently linked
CC   to e.g., keyhole limpet haemocyanin, BSA, ovalbumin, poly-L-lysine,
CC   or VP6 bovine rotavirus protein. An adjuvant may be included.
SQ   Sequence   1357 BP;    449 A;    257 C;    267 G;    384 T;

Initial Score    =     14  Optimized Score  =     27  Significance =  5.51
Residue Identity =    18%  Matches          =     31  Mismatches   =   124
Gaps             =     10  Conservative Substitutions               =     0
Translation Frame=      1

             X       10          20        30        40            50
             QDCSFQHSPISSDFAVK--IRELSDYLLQDYPVTVASNLQDEEL----CGGLWRLVLAQ--R
                  |    |      |  |    |             |    |       |     ||
   LNQTPXESCQPLNSKEXILIIRRNTXKTGICKIEDRGQVSLFINQTFFLIQHHLHXIDHNPLMIIXWAQCGX
 120       130       140       150       160       170       180       190

     60        70        80        90        100       110       120
   WMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLR--FVQTNISRLLQETSEQLVALKPWITRQNFS
      |      ||          |         ||  | ||      || | || |       | ||
   TQDRKFKSLDLTTHVLLTHQPIYNNLSILCHSDECXLQLRXLFYQTRKGLVFQEXSIQLTAQLHGFSTQXFS
        200       210       220       230       240       250       260

      130       140       150    X
   RCLELQCQPDSSTLPPPWSPRPLEATAPTAP
      |              ||
   GRITLKWSFYXMDRXXTLIKQDLELSXLEILILLDYHSSXX
       270       280       290  X    300
```

15. US-08-162-407-6 (1-235)
    N90612 CD19 cDNA.

```
ID   N90612 standard; cDNA; 1921 BP.
AC   N90612;
DT   20-DEC-1989  (first entry).
DE   CD19 cDNA.
KW   Tyrosine suppressor gene; ASV LTR; human cytomegalovirus AD169 enhancer;
KW   HIV box; immunoselection; immune deficiency diseases;  vasculitis;
KW   systemic lupus erythematosus; rheumatoid arthritis; neoplasms;
PN   EP-330191-A.
PD   30-AUG-1989.
PF   23-FEB-1989; 103127.
PR   25-FEB-1988; US-160416.
PA   (GEHO) Gen. Hospital Corp.
PI   Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo CP,
PI   Simmons D, Stamenkovic I, Stengelin S;
DR   WPI; 89-250302/35.
PT   Rapid immuno:selection cloning - used to clone genes encoding
PT   cell surface antigens associated with mammalian T lymphocytes.
PS   Disclosure; fig. 12; 69pp; English.
CC   CD19 encodes a cell surface antigen involved in cell mediated immunity.
CC   This DNA can be expressed in a vector which transforms COS cells. The
CC   vector can isolate any protein, and clones are easy to manipulate.
SQ   Sequence   1921 BP;   402 A;   566 C;   567 G;   386 T;

Initial Score    =     14  Optimized Score  =     31  Significance =  5.51
Residue Identity =    21%  Matches          =     39  Mismatches   =   115
Gaps             =     26  Conservative Substitutions               =     0
```

Translation Frame= 1

```
             X       10        20                30        40        50
             QDCSFQHSPISSDFAVKIREL--------SDYLLQDYPVTVASNLQDEELCGGLWRLVLAQR
              |    ||        |    |                   |   |  ||     |||
   GRXRCAAVPQGDLRWPHSAADLVSGVPAXTLLKTQPGAARPGNPHEAPGHLAFHLQRLSTDGGLLPVPAGAP
      40     X  50        60        70        80        90        100

          60        70        80        90       100       110       120
   WMERLKTVAGSKMQGLLERVNTEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLV-ALKPWITRQNFSR
     | |     |   |    |  |          |    || |    |   |     |    |       |
   LXEGLAAWLDSQCGGQRGAVPVECFGPRWPGLWPEEQVLRGPQ-LPFREAHEPQAVCVGQRPPXDLGGRASV
   110       120       130       140       150       160       170

      130         140         150               X
   C-LELQCQPDSSTLP--PPW--SPRPLEAT-----------APTAP
   |  | |  |     |   ||      |  |             | ||
   CPTEGQPEPEPQPGPHHGPWLHTLAVLWGTPXLCVQGPPLLDPCAPQGAXVIAEPR
 180       190       200       210       220   X   230
```

maryh@stic **stdin**

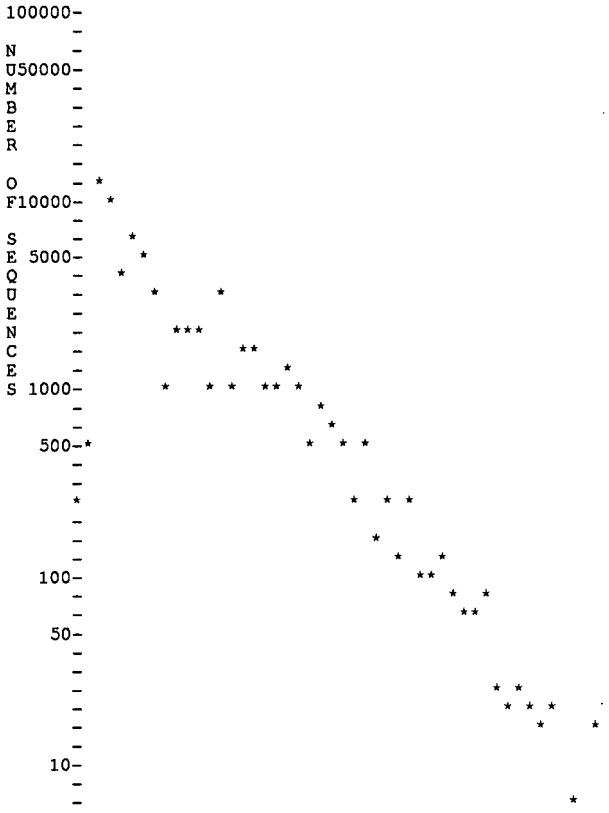
NeWSprinter20

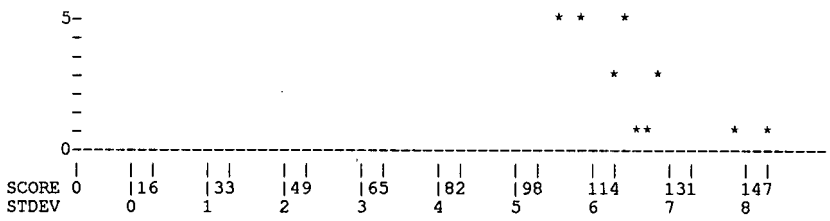
Thu Apr 06 10:25:06 1995

NeWSprint 2.5 Rev B
Openwin library 3
NeWSprint interpreter 210.0

NeWSprint 2.5

```
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file trans_nsq.res made by root on Thu 6 Apr 95 9:58:03-PDT.


Query sequence being compared:US-08-162-407-6 (1-705)
Number of sequences searched:                57621
Number of scores above cutoff:                4419

     Results of the initial comparison of US-08-162-407-6 (1-705) with:
   Data bank : N-GeneSeq 17, all entries
```





PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	19	14	15.88

Times:	CPU	Total Elapsed
	00:03:44.07	00:03:45.00

Number of residues:	24347505
Number of sequences searched:	57621
Number of scores above cutoff:	4419

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 8 standard deviations above mean ****					
1. N91235	DNA sequence of prepr0spasmol	1368	147	368	8.06	0
	**** 7 standard deviations above mean ****					
2. Q43249	Sequence of AtT20 murine proh	2012	138	373	7.49	0
	**** 6 standard deviations above mean ****					
3. Q45334	Human TLE-2 gene.	2271	122	368	6.48	0
4. Q62176	Human TLE-2 gene.	2271	122	368	6.48	0
5. Q11127	Sequence encoding P450SU2 and	1998	121	365	6.42	0
6. Q27764	Adrenodoxin reductase gene (p	5089	118	351	6.23	0
7. Q56917	Glutamic acid receptor.	3717	117	369	6.17	0
8. N60554	Sequence encoding bovine male	5783	117	366	6.17	0
9. Q06233	Bovine male-specific insert o	4414	116	367	6.11	0
10. Q02830	cDNA to HIV-2 RNA.	9360	116	348	6.11	0
	**** 5 standard deviations above mean ****					

11.	Q46249	Construct EC2L (Contains catR	8528	114	360	5.98	0
12.	Q31999	Encodes eimeria antigen Eam45	1631	113	372	5.92	0
13.	Q40697	B. burgdorferi strain Ip90 Os	1005	110	346	5.73	0
14.	Q14687	Factor XIII subunit a.	2350	110	261	5.73	0
15.	N82401	A and A' subunits of human Fa	2350	110	262	5.73	0
16.	Q25893	Human Factor XIII coding sequ	3905	110	261	5.73	0
17.	N70461	Sequence encoding human facto	3905	110	259	5.73	0
18.	Q58719	Human gamma-1 chain second me	838	109	230	5.67	0
19.	Q55783	Segment of human gamma-1 chai	838	109	230	5.67	0
20.	Q58990	Mutant human creatine kinase	1568	109	362	5.67	0

1. US-08-162-407-6 (1-705)
   N91235 DNA sequence of prepropasmolysin.

```
 ID   N91235 standard; DNA; 1368 BP.
 AC   N91235;
 DT   07-FEB-1990  (first entry)
 DE   DNA sequence of prepropasmolysin.
 KW   Prepropasmolysin; Xenopus laevis.
 OS   Xenopus laevis
 FH   Key             Location/Qualifiers
 FT   CDS             28..1230
 FT   /*tag= a
 PN   DE3808456-A.
 PD   28-SEP-1989.
 PF   14-MAR-1988; 808456.
 PR   14-MAR-1988; DE-808456.
 PA   (PLAC) Max-Plank Ges. Wissensch.
 PI   Weickmann H, Fincke K, Weickmann F, Huber B, Liska H, Prechtel J,
 DR   WPI; 89-286097/40. P-PSDB; P91941.
 PT   New polypeptide preprospasmolysin and its fragments- derived from Xenopus
 PT   laevis, with spasmolytic and neurotransmitter activities.
 PS   Disclosure; Fig. 2; 8pp; german.
 CC   The DNA or mRNA corresp. to the peptides are derived from the skin of
 CC   Xenopus laevis. The DNA sequence is inserted into a vector and used to
 CC   transform eukaryotic cells, esp. COS cells, or the DNA is converted into
 CC   mRNA and microinjected into eukaryotic cells, esp. X. laevis oocytes.
 SQ   Sequence   1368 BP;   511 A;   332 C;   234 G;   291 T;

Initial Score    =    147  Optimized Score  =    368  Significance =  8.06
Residue Identity =    26%  Matches          =    338  Mismatches   =   340
Gaps             =     81  Conservative Substitutions               =     0

            X       10        20        30        40        50        60
            ATGACNGTNYTNGCNCCNGCNTGGWSNCCNACNACNTAYYTNYTNY-TNYTNYTNYTNYTNW
            ||| | |  ||     || |  |   |   | || |  |  |  || | |  | |  |  |
   GAGCAGTAGGATG-CTGCTTTGATTCCAGTATCCTTAAC-ACTAAATGGTGCTTCTATAATGCAACAGC--A
   160       170       180       190       200       210       220

         70          80          90         100         110         120
   SNWSNGGNY---TNWSNGGNACNCA----RGAYTGYWSNTTYCARCAYWS---NCCNATNWSNWSNGAYTTY
   | ||    |     ||  | |  ||    ||||  ||| |   || ||||    |  ||  | |   ||  |
   GGTCCAATCAAGAAACTCGAATGCAGTGGGGATCCTAC-TAAGAGAATAGACTGCGGATTCCCAAGAATAAC
     230       240       250       260       270       280       290

      130       140         150       160       170         180
   GCNGTNAARATNMGNGARYTNWS-NGA--YTAYYT--NYTNCARGAYTAYCCNG---TNACNGTNGCNWSNA
      |  | |||  |    || ||  ||  || |||   || ||| ||| ||| |   | |  |  || ||
   AGAGAAACAAT--GCATTCTAAGAGGATGCTGTTTTGATTCCAGTATT-TCCGGTGTTAAATGGTGCTACGC
```



```
    300         310         320         330         340         350         360

  190       200         210         220         230         240         250
  AYYTNCARGAYGARGARYT-NTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNY--
  || | |||    || ||||   ||    | |     |  ||    ||    ||||    ||  |||  |
  ACGTACAGTTATAACAACTCCAGCACCAGAT----ACAACTACAGCTTCAACAACTGCAGAAACAACTACAG
    370       380       390         400       410       420       430

   260       270         280       290       300       310       320
  TNAARACNGTNGCNGGNWSNA----ARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTN
  |   |||     | |  || |    || | |||   || |   |||   |  |||     |   ||| |
  TTCCAACAACTCCAGAAACAACTACAGTTCCAACAACTCCAGAAACAACT--ACAGTTCCAACAACTCCAGA
      440       450       460       470       480       490       500

     330       340       350       360       370       380       390
  ACNAARTGYGCNTTYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCAR-ACNA-AYATNWSNMGNYTNYTNC
  |  || | |  |||||||  |  |  |     || |  |||   |||  | | | |  || |  |  || |
  AACAACTAC-AGTTCCAACAACTCCAGAAACAACTACAGTTCCAACAACTCCAGAAACAACTA--CAGTTCC
      510       520       530       540       550       560       570

     400       410       420       430       440         450
  ARGAR-ACNWSNGARCARYT--NGTNGCNYTNAARCCNTGGATNACNMGNCAR----AAYTTYWSNMGNTGY
  || ||  | ||  |||||||   ||  |    |  ||    |  || |  |||    |||  ||| ||   |
  AACAACTCCAG-AAACAACTACAGTTCCAACAACTCCAGAAACAACTA--CAGTTCCAACAACTCCAGAAAC
       580       590       600       610       620       630       640

 460       470         480       490       500       510       520
  YTNGARYTNCARTGYCAR-----CCNGAYWSNWSNAC-NYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGA-
      |||| |||| ||||     || || || || ||  || |  |  |    ||  |  |  |  ||  |
  ----AACTACAGTTCCAACAACTCCAGAAACAACTACAGTTCCAACAACTCCAGAAACAACTACAGTTCCAA
           650       660       670       680       690       700

     530       540       550       560       570       580       590
  -RGCNACNGCNCCNACNGC-NCCNCARCCNCCNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNYTN
   | |  | |   | ||  |  |  ||||  |  |     |  || |  || |  |   |        |  ||
  CAACTCCAGAAACAACTACAGCTTCAACAACTGCAGAAACAACTACAGTTCCAACAACTCCAGAAACAACTA
710       720       730       740       750       760       770       780

    600       610       620       630       640       650       660
  YTNGCNGCNGC-NTGGTGYYTNC-AYTGGCARMGNACNMGNMGNMGNA-CNCCNMGNCCNGGNGARCARGTN
  |  |   |  |          ||  | || |  |     || |  |  |  | | || |  || |   ||||| |
  C-AGAACCGACTACAACACCAACAACAGATACTACACCACCAACACTACCACCAACACCAG--AAACAACTA
         790       800       810       820       830       840       850

       670         680       690       700.    X
  C---CNCCNGTNCCNW---SNCCNCARGAYYTNYTNYTNGTNGARC-AY
  |     |  | |  |      |  ||| | |  || |       ||| ||
  CAGAAACAACTACAGAAACAACTACAGAAACAACTACAGAAACAACTACAGAAACAACT
         860       870       880       890       900
```

2. US-08-162-407-6 (1-705)
   Q43249 Sequence of AtT20 murine prohormone convertase 2 c

```
 ID   Q43249 standard; cDNA; 2012 BP.
 AC   Q43249;
 DT   24-SEP-1993  (first entry)
 DE   Sequence of AtT20 murine prohormone convertase 2 cDNA.
 KW   Prohormone convertase; enzyme; cleavage; ss.
 OS   Mus musculus.
```

```
FH   Key             Location/Qualifiers
FT   CDS             6..2012
FT   /*tag=  a
PN   WO9311247-A.
PD   10-JUN-1993.
PF   04-DEC-1992; U10621.
PR   06-DEC-1991; US-803631.
PR   22-MAY-1992; US-887265.
PA   (GETH ) GENENTECH INC.
PI   Gorman CM,  Groskreutz DJ,  Marriott D;
DR   WPI; 93-197065/24.
DR   P-PSDB; R37618.
PT   Heterologous polypeptide factor prepn. - by introducing into
PT   polypeptide factor dependent host cell nucleic acid, and then
PT   culturing host cell, etc.
PS   Disclosure; Fig 2; 134pp; English.
CC   The inventors describe the production of mammalian cells expressing
CC   prohormone convertase which facilitates the processing of prohormone
CC   precursors to active hormones. More specifically the cleavage site
CC   is the prohormone convertase cleavage site given in R37632-35. The
CC   mouse pituitary tumour cell line, AtT20, was used as the source for
CC   candidate prohormone convertase mRNA's from which cDNA was generated.
CC   Amplification using PCR of candidate prohormone convertase targets
CC   was carried out. The primer sequences were based on the conserved
CC   aspartate and serine catalytic residues of KEX2, PC2, Proprotease B,
CC   and subtilisin BPN. The forward primer used was Q43250 and the
CC   reverse primer was Q43251. A cDNA pool was prepared as a source of
CC   template for the cloning of full length sequences of the candidate PC
CC   enzymes. Mouse AtT-20 total RNA was denatured and added to a first
CC   strand cDNA synthesis reaction containing adapter primer Q43252.
CC   Aliquots of the cDNA pool were used for amplification of unknown 3'
CC   and 5' regions by Q43253  and 5' RACE primer Q43254 and the inner
CC   adapter Ri PCR primer Q43255.
SQ   Sequence   2012 BP;    520 A;    516 C;    565 G;    411 T;
```

Initial Score	=	138	Optimized Score	=	373	Significance = 7.49
Residue Identity	=	28%	Matches	=	350	Mismatches = 313
Gaps	=	88	Conservative Substitutions		=	0

```
            X       10          20          30          40          50
            ATGACNGTNYTNGCNC---CNGCNTG--GWSNCCNACNACNTAYYTNYTNYTNYTNYTN
            || ||  | ||  | |     || ||   || || |  |   |     |  |  || |  |
   GAAGGCCTGTATCAC--TTTTATCACAATGGGCTTGCAAAGGCCAA--AAGAAGACGCAGC--CTACACC--
    270       280         290         300         310         320         330

     60        70        80        90        100       110       120
   YTNWSNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTN
     | ||  |  | |  || |  || | ||  |  || |    || |||  | |  || || |||||  | ||
   ATAAGCGGCAGCTAGAGAGAGACCCCAGGATAAAGAT---GGCGCTG--CAACAAGAAG-GATTTGACCGT-
           340       350       360         370         380         390

  130       140         150         160         170         180         190
    AARATNMGNGARYTNWSNGAYTAYYTNYTNCA---RGAYTAYCCNG--TNA--CNGTNGCNWSNAAYY--TN
    ||||  || | ||  || ||| |||  |   |   | || | |  |  | |  |  |    || || |  |
    AAAAAGAGAGGGTACAGGGAC-ATCAATGAGATTGACATCAACATGAATGATCCTCTCTTTACAAAGCAATG
      400       410         420         430         440         450         460

         200       210         220         230         240         250         260
    CARGAYGARGARYTNTGYGGNGGNYTNTGG-MGNYTNGTNYTNG-CNCARMGNTGGATGGARMGNYTNAARA
      |  ||   |||  || |   | |  ||| |  |  |  || | |   ||  |||  |  ||| ||   |
```



```
GTACCTGTTCAACACTGGGCAAGCCGATGGAACTCCTGGGCTAGACTTGAACGTGGCCG--AAGCCTGGGAG
 470       480       490       500       510       520       530

     270       280       290       300       310       320       330
CNGTNGCNGGNWSNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYT-TYGTNAC-NAARTG
| |       || |||| | || | |  || | ||   |  ||   | |||  ||| ||   ||  | |
CTGGGATACACAGGAAAA-GGAGTGACCATT-GGAATTATGGATGATGGAATTGACTATCTCCACCCAGACC
 540       550       560       570       580       590       600

     340       350       360       370       380       390       400
YGCNTTYCARCCNCCNCCNWSNTGYYTNMGNTTYG--TNCARACNAAYATNWSNMGNYTNYTNCARGARACN
||   | ||||  |  |  || |||   || | ||  | |||   | ||  || |  || |  | | |  |
TGGCCTACAACTACAACGCTGATGC--AAGTTATGACTTCAG--CAGCA-ATGACCCCTAC--CCATACCCT
 610       620       630       640       650       660       670

      410       420       430       440               450       460
WSNGA-RCA-RYTNGTNGCNYTNAARCCNTGGATNACNMG----NCA----RAAYTTYWSNMGNTGYYTNGA
 |  | ||| ||  | |  |   |||| ||||  || ||     ||    || |||||  |  |||   |
CGATACACAGATGACTGGTTCAACAGCCATGGA--ACTAGGTGTGCAGGAGAAGTTTCTGCTGCAGCC---A
      680       690       700       710       720       730

   470       480       490       500       510       520       530
RYTNCARTGYCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGC--NAC-NGC
||  |||| || |   || ||  |    |   | | ||  |||| |  ||    |  || ||   ||  ||
GCAACAATATCTG-TGGAGTCGGCGTAGCATACAACTCCAAGGTGGC--AGGGATCCGGATGCTGGACCAGC
740       750       760       770       780       790       800

        540       550       560       570       580       590
NC----CNACNGCNCCNCARCCNCCNYTNYTNYTNYTNYTNY-TNYTNC--CNGTNGGNYTNYTNYTNYTNG
 |      || |  |  || |     || || |  |  |  | | |  |  | |  |  | || | |  |
CCTTTATGACAG-ACATCATCGAAGCCTCCTCCATCAGCCACATGCCTCAACTGATCGACATCTACAGT--G
810       820       830       840       850       860       870

600       610       620       630       640       650       660
CNGCNGCNTGGTGYYTNCAYTGGCAR-MGNACNM--GNMGNMGNACNCCNMG-NCCNGGNGARCARGTNCCN
|    || ||| |||  ||| | |||  | |  |  |  |  |  | || ||  |         ||||
C--AAGC-TGGGGCC-CCACAGACAATGGGAAGACGGTTGATGGGCCCCGAGAGCTCACACTCCAGGCCATG
   880       890       900       910       920       930       940

 670       680       690       700    X
CCNG-TNCCNWSNCCNCARGAY-YTNYTNYTNGTNGARCAY
 | | |  | ||  |  || | |      |      |||||
GCTGATGGCGTGAACAAGGGCCGTGGGGGCAAAGGCAGCATCTATGTGTGG
   950       960       970       980    X  990
```

3. US-08-162-407-6 (1-705)
   Q45334 Human TLE-2 gene.

```
ID   Q45334 standard; cDNA; 2271 BP.
AC   Q45334;
DT   10-SEP-1994  (first entry)
DE   Human TLE-2 gene.
KW   TLE-2; transducin-like enhancer of split protein; cell fate;
KW   differentiation; cervix cancer; breast cancer; psoriasis; baldness;
KW   ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             26..2257
FT   /*tag=  a
```

```
PN   WO9407522-A.
PD   14-APR-1994.
PF   30-SEP-1993; U09339.
PR   30-SEP-1992; US-954813.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-tsakonas S,  Stifani S;
DR   WPI; 94-135221/16.
DR   P-PSDB; R51110.
PT   Transducin-like enhancer or split proteins and nucleic acids- -
PT   are for treatment of disorders of cell fate or differentiation
PT   e.g. cervical cancer, breast cancer, psoriasis, baldness etc.
PS   Disclosure; Page 91-94; 147pp; English.
CC   The nucleotide and deduced aa sequences of human transducin-like
CC   enhancer of split proteins TLE-1 (Q45333, R51109), TLE-2 (Q45334,
CC   R51110), TLE-3 (Q45335, R51111) and TLE-4 (Q45336, R52953) were
CC   determined.  The aa sequences were compared with that of
CC   Drosophila E(spl) m9/10 (R52955).  Comparison of the WD-40 domains
CC   of these proteins defined the consensus residues shown in R52954.
CC   The CcN motifs of the proteins were compared with those of SV40 T
CC   antigen, human c-myc, human p53, human A-myb and dorsal protein with
CC   respect to nuclear localization site, and casein-kinase and cd2-
CC   kinase phosphorylation sites (sequences R52956-70).
SQ   Sequence   2271 BP;    437 A;    751 C;    672 G;    411 T;

Initial Score     =    122  Optimized Score  =    368  Significance =  6.48
Residue Identity  =    27%  Matches          =    338  Mismatches   =   333
Gaps              =     75  Conservative Substitutions               =     0

            X       10        20          30        40        50        60
            ATGACNGTNYTNGCNCCNGC--NTGGWSNCCNACNACNTAYYTNYTNYTNYTNYTNYTN
              ||||        ||  | ||    ||||  | |   | ||        |        | |
   TATCCCCTTCCTGAC-CCAGGAGCATCAGCAGCAGGTGCTCCAGGCCGTAGAACGCGCCAAGCAGGTCACCG
   300       310         320       330       340       350       360

           70        80        90        100       110       120
   WSNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCAR-CAYWSNCCNATNW-SNWSNGAYTTYG-CNGTN
   ||  |  | |  || | |   |   || | ||||  ||||| | ||| || |  | | ||   || || |  |
   TGGGGGAGCTGAACAGCCTCATCGGGCAGCAG-CTCCAGCCGCTGTCCCACCACGCACCCCCTGTGCCCCTG
  370       380       390       400       410       420       430

  130       140       150       160           170       180       190
   AARATNMG-NGARYTNWSNGAYTAYYTNYTNCARGAYTAY----CCNGTNACNGTNGCNWSNAAYYTNCARG
   |     ||   |||   | | |  |   | |||  ||||    |   |  | |  |  |   |||    |
   ACCCCCCGCCCAGC--CGGGCTGGTGGGCGGCAGTGCTACGGGGCTGCTTGCTCTGTCTGGAGCCCTGGCTG
  440       450       460       470       480       490       500

     200         210       220       230       240       250       260
   AYGARG---ARYTNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARM-GNYTNAARACN
    | |||   ||||  |||| || || |  ||    |  |  ||   ||| |||| |  || | |  |||
   CCCAGGCTCAGCT-GGCGGCGG-CTGTCAAG-GAGGACCGTGC-GGGCG-TGGAGGCCGAGGGGTCCAGAGT
  510       520       530         540       550       560       570

      270       280       290       300       310       320       330
   GTNGCNGGNWSNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCN
   |  |   | |  ||  | || | |  ||    ||      | | ||||    | ||||     || | ||
   GGAGAGAGCCCCGAGCAGGAGTG-CATCTCCCTCG----CCCCCTGAGA---GTCTCGTGGAGGAG-GAGCG
    580       590       600       610         620         630

    340       350       360       370       380       390       400
   TTYCARCCNCCNCCNWSNTGYY--TNMGNTTYGTNCARACNAAYATNWSNMGNY--TNYTNCARGARA-CNW
```



```
     | ||   ||    || || |    ||    | ||||  |  |  || | |    |  | |||| | |
   ACCGAGTGGCCCTGGTGGTGGCGGGAAGCAGAGAGCAGATGAGAAGGAGCCATCAGGACCTTATGAAAGCGA
640       650       660       670       680       690       700       710

       410       420       430       440       450       460       470
   SNGARCARYTNGTNGCNYTNAARC-CNTGGATNACNM-GNCARAAYTTYWSNMGNTGYYTNG---ARYTNCA
   |  ||  |   || |  |  || |   |||   || | | | || | |||| ||   ||  |   ||| |
   CGAAGACAAGAGT-GATTACAATCTGGTGGTGGACGAGGACCAACCCTCAG-AGCCCCCCAGCCCGGCTACC
         720       730       740       750       760       770       780

         480       490       500       510       520       530       540
   RTGYCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGC
   |  ||   | || ||  | || |  |  |   | ||  | |  ||    || | ||  || | ||  | |
   ACCCCCTGCGGAAAG--GTACCCATCTGCATTCCTG--CCCGTCGGGACCTGGTGG--ACAG-TCCAGC--C
         790       800       810       820       830       840

       550       560       570       580       590       600
   NCCNCARCCNC--CNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNYT------NYTNGCNG-CNGC
    ||   ||| |   || | |  |  | ||       |   |  ||  | ||       ||  | | | ||
   TCCTTGGCCTCTAGCTTGCGGTCACCGCTGCCTAGAGCCAAGGAGCTCATCCTGAATGACCTTCCCGCCAGC
     850       860       870       880       890       900       910

     610           620       630       640       650       660
   -NTGGTGYYTNCA----YTG-GCARMGNACNMGNMGNMGNACNC--CNMGNCCNGGNGARCARGTNCCNCCN
     |  ||||| ||    ||| |   |   | || |  || || |  | |  || || |  ||| | |  ||
   ACTCCTGCCTCCAAATCCTGTGACTCCTCCCCGCCCCAGGACGCTTCCACCCCCGG-GCCCAGCT-CGGCCA
    920       930       940       950       960       970       980

670       680       690       700    X
  GTNCCNWSNCCNCARGAYYTNYTNYTNGTNGARCAY
  ||  | ||  | |||  ||| |  |      |||||
  GTCACCTCTGC-CAG--CTTGCGCTCAAGCCAGCACCTTCCACGGA
   990       1000        1010      1020
```

4. US-08-162-407-6 (1-705)
   Q62176 Human TLE-2 gene.

```
ID   Q62176 standard; cDNA; 2271 BP.
AC   Q62176;
DT   10-SEP-1994  (first entry)
DE   Human TLE-2 gene.
KW   TLE-2; transducin-like enhancer of split protein; differentiation;
KW   protein transport; cervix cancer; dysplasia; malignancy;
KW   ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             26..2257
FT   /*tag=  a
PN   WO9408037-A.
PD   14-APR-1994.
PF   30-SEP-1993; U09333.
PR   30-SEP-1992; US-955011.
PA   (MEDI-) MEDICAL RES COUNCIL.
PA   (UYYA ) UNIV YALE.
PI   Artavanis-tsakonas S,  Hill RE,  Redhead NJ,  Stifani S;
DR   WPI; 94-135597/16.
DR   P-PSDB; R51477.
PT   New human transducin-like enhancers of split protein - and
PT   associated multi-protein complexes, chimeric proteins,
```

```
PT   antibodies, nucleic acid, etc., involved in nuclear-cytoplasmic
PT   protein transport
PS   Disclosure; Page 58-61; 112pp; English.
CC   The nucleotide and deduced aa sequences of human transducin-like
CC   enhancer of split proteins TLE-1 (Q62175, R51476), TLE-2 (Q62176,
CC   R51477), TLE-3 (Q62177, R51478) and TLE-4 (Q62178, R51479) were
CC   determined.  The aa sequences were compared with that of Drosophila
CC   E(spl)m9/10 protein (R51481).  Comparison of the WD-40 domains of
CC   these proteins defined the consensus residues shown in R51480.  The
CC   CcN motifs of the proteins were compared with those of the SV40 T
CC   antigen, human c-myc, human p53, human A-myb and dorsal proteins
CC   with respect to nuclear localization site, and casein-kinase II and
CC   cdc2-kinase phosphorylation sites (sequences R51482-96).  TLE can be
CC   used to treat or diagnose (pre)neoplastic conditions, or
CC   to study cell differentiation.
SQ   Sequence   2271 BP;    437 A;    751 C;    672 G;    411 T;

Initial Score     =    122  Optimized Score  =    368  Significance =  6.48
Residue Identity  =     27% Matches          =    338  Mismatches   =   333
Gaps              =      75 Conservative Substitutions              =     0

            X       10        20        30        40        50        60
            ATGACNGTNYTNGCNCCNGC--NTGGWSNCCNACNACNTAYYTNYTNYTNYTNYTNYTN
             ||||       ||  | ||    ||||  | |   | ||      |          |  |
   TATCCCCTTCCTGAC-CCAGGAGCATCAGCAGCAGGTGCTCCAGGCCGTAGAACGCGCCAAGCAGGTCACCG
   300       310       320       330       340       350       360

           70        80        90        100       110       120
   WSNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCAR-CAYWSNCCNATNW-SNWSNGAYTTYG-CNGTN
   ||  |  | |  || |   |   || | ||||  ||||| | ||| || |  | | ||   || || |  |
   TGGGGGAGCTGAACAGCCTCATCGGGCAGCAG-CTCCAGCCGCTGTCCCACCACGCACCCCCTGTGCCCCTG
  370       380       390       400       410       420       430

 130       140       150       160           170       180       190
   AARATNMG-NGARYTNWSNGAYTAYYTNYTNCARGAYTAY----CCNGTNACNGTNGCNWSNAAYYTNCARG
   |     ||   |||   | | |  |   |  |||  ||||    |   |  |  |  |  |  |  |||    |
   ACCCCCCGCCCAGC--CGGGCTGGTGGGCGGCAGTGCTACGGGGCTGCTTGCTCTGTCTGGAGCCCTGGCTG
 440       450       460       470       480       490       500

    200       210       220       230       240       250       260
   AYGARG---ARYTNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARM-GNYTNAARACN
    | |||   ||||  |||| || || |  ||    |  |  ||   ||| |||| |  || |  |  |||
   CCCAGGCTCAGCT-GGCGGCGG-CTGTCAAG-GAGGACCGTGC-GGGCG-TGGAGGCCGAGGGGTCCAGAGT
 510       520       530       540       550       560       570

     270       280       290       300       310       320       330
   GTNGCNGGNWSNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCN
   |  |   | |  ||  | || | |  ||    ||       | | ||||    | ||||     || | ||
   GGAGAGAGCCCCGAGCAGGAGTG-CATCTCCCTCG----CCCCCTGAGA---GTCTCGTGGAGGAG-GAGCG
    580       590       600       610           620           630

    340       350       360       370       380       390       400
   TTYCARCCNCCNCCNWSNTGYY--TNMGNTTYGTNCARACNAAYATNWSNMGNY--TNYTNCARGARA-CNW
     | ||   ||    || || |    ||    |  |||| |  |  || | |    |   | |||| | |
   ACCGAGTGGCCCTGGTGGTGGCGGGAAGCAGAGAGCAGATGAGAAGGAGCCATCAGGACCTTATGAAAGCGA
 640       650       660       670       680       690       700       710

      410       420       430       440       450       460       470
   SNGARCARYTNGTNGCNYTNAARC-CNTGGATNACNM-GNCARAAYTTYWSNMGNTGYYTNG---ARYTNCA
   |  ||  |   || |  |  || |   |||   || | | | || | |||| ||   ||  |    ||| |
```



```
   CGAAGACAAGAGT-GATTACAATCTGGTGGTGGACGAGGACCAACCCTCAG-AGCCCCCCAGCCCGGCTACC
         720       730       740       750       760       770       780

          480       490       500       510       520       530       540
   RTGYCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGC
   |  ||    | || ||  | || |  |  |    | ||  | |  ||     || | ||  || |  ||  | |
   ACCCCCTGCGGAAAG--GTACCCATCTGCATTCCTG--CCCGTCGGGACCTGGTGG--ACAG-TCCAGC--C
         790       800       810       820       830       840

        550       560       570       580       590       600
   NCCNCARCCNC--CNYTNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNYT------NYTNGCNG-CNGC
    ||   ||| |    || |  |  |  || |       |    |  ||  | ||       ||  | | | ||
   TCCTTGGCCTCTAGCTTGCGGTCACCGCTGCCTAGAGCCAAGGAGCTCATCCTGAATGACCTTCCCGCCAGC
      850       860       870       880       890       900       910

     610           620       630       640       650       660
   -NTGGTGYYTNCA----YTG-GCARMGNACNMGNMGNMGNACNC--CNMGNCCNGGNGARCARGTNCCNCCN
     |  ||||| ||    ||| |    |  | || |  || || |  | |  || || |  ||| | |  ||
   ACTCCTGCCTCCAAATCCTGTGACTCCTCCCCGCCCCAGGACGCTTCCACCCCCGG-GCCCAGCT-CGGCCA
    920       930       940       950       960       970       980

 670       680       690       700    X
   GTNCCNWSNCCNCARGAYYTNYTNYTNGTNGARCAY
   ||  | ||  | |||  ||| |  |      |||||
   GTCACCTCTGC-CAG--CTTGCGCTCAAGCCAGCACCTTCCACGGA
    990       1000      1010      1020


5. US-08-162-407-6 (1-705)
   Q11127       Sequence encoding P450SU2 and FeS-A.

 ID   Q11127 standard; DNA; 1998 BP.
 AC   Q11127;
 DT   05-JUN-1991  (first entry)
 DE   Sequence encoding P450SU2 and FeS-A.
 KW   Cytochrome P450; P450SU1; P450SU2; herbicide resistance; ds.
 OS   Streptomyces griseolus.
 FH   Key             Location/Qualifiers
 FT   misc_RNA        195..1403
 FT   /*tag=  a
 FT   /product= P450SU2
 FT   misc_RNA        1452..1643
 FT   /*tag=  b
 FT   /product= FeS-A protein
 PN   WO9103561-A.
 PD   21-MAR-1991.
 PF   27-AUG-1990; U04785.
 PR   11-SEP-1989; US-405605.
 PR   12-JAN-1990; US-464499.
 PR   23-AUG-1990; US-569781.
 PA   (DUPO ) DU PONT DE NEMOURS CO.
 PI   Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;
 PI   O'Keefe DP, Omer CA, Romesser JA;
 DR   WPI; 91-102077/14.
 DR   P-PSDB; R11350, R11804.
 PT   DNA encoding cytochrome P450 enzymes - and electron donating
 PT   iron-sulphur proteins, used to confer herbicide resistance to
 PT   plants and microorganisms
 PS   Claim 15; page 158; 224pp; English.
 CC   This sequence contains the coding information for (a) the cytochrome
```

```
CC   P450 enzyme, P450SU2; and (b) the electron donor iron sulphur
CC   protein FeS-A.  It is contained in a recombinant plasmid which is
CC   used to transform host Streptomyces species which in turn are used
CC   to coat a plant seed to transform the plant.  The resultant trans-
CC   formants are resistant to herbicides.
CC   See also Q11126.
SQ   Sequence   1998 BP;    272 A;    839 C;    620 G;    267 T;

Initial Score    =    121  Optimized Score  =    365  Significance =  6.42
Residue Identity =     28%  Matches         =    331  Mismatches   =   346
Gaps             =     65  Conservative Substitutions              =     0

            X       10        20        30        40        50        60
            ATGACNGTNYTNGCNCCNGCNTGGWSNCC-NACNACNTAYYTNYTNYTNYTNYTNYTNW
            | | | |       |  | || ||  | ||  |  ||  ||| |  |  |  |  |  |
    TAATTAGCTCACG-CGGAAGGACCGGCCGCATGACGACCGCAGAACGCACCGCTCCCCCCGACGCCCTCACC
      170     X 180       190       200       210       220       230

           70        80        90       100       110       120       130
    SNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAY-WSNCCNATNWSNWSNGAYTTYG-CNGTNAA
    |  | || |  ||  |  | |  | |||| | ||| | | |  | || |  || || ||    | | |
    GTCCCGGCC--AGCCGCGCCCCCGGCTGCCCCTTCGACCCCGCGCCCGACGTC-ACCGAGGCGGCCCGCACC
     240       250       260       270       280       290       300

          140       150       160       170       180       190       200
    RATNMGNGARYTNWSNGAYTAYYTNYTNCA-RGAYTAYCCNG-TNACNGTNGCNWSNAAYYTNCARGAYG-A
    ||  ||   ||  | | | ||| |   | || |  ||| | |  |  | | ||  ||  |||||||
    GAACCGGTCAC-CCGGGCCACCCTCTGGGACGGCTCCTCCTGCTGGC--TGGTGACGCGCCATCAGGACGTC
      310       320       330       340       350       360       370

             210       220       230           240       250       260
    RGARY--TNTGYGGNGGNYTNTGGMGNYT-NGTNY--TNGCNCA--RMGNTGGATGGARMGNYTNAARACNG
     | ||  |   ||| |  |    ||| ||  |  |    || ||  ||   || |    |  ||   ||| |
    CGCGCGGTCCTCGGCGACC---CGCGCTTCAGCGCCGACGCCCACCGCACCGGCT--TCCCCTTCCTGACCG
       380       390           400       410       420           430       440

      270       280       290       300       310       320       330
    TNGCNGGNWSNAARAT--GCARGGNYTNYTNGARM-GNGTNAAYACNGA-RATNCAYTTYGTNACNAARTGY
      |  ||  |  ||||   | |   |  |  || |     |   ||  || ||  |    |    |  | |||
    CCGGCGGCCGCGAGATCATCGGCACCAACCCGACCTTCCTGCGCATGGACGACCCGGAGCACGCCCGACTGC
          450       460       470       480       490       500       510

           340       350       360       370       380       390       400
    GC---NTTYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACN
    ||    | |  ||| ||  | || | | | |     ||  || |  ||    | ||     | |||
    GCCGGATGCTCACCGCCGACTTCAT-CGTCAAGAAGGTCGAGGC-GATGCGCCCCG--AGGTGCAGCGCCTC
        520       530       540       550       560       570       580

         410       420       430       440       450       460       470
    WSNGARCARYTNGTNGCNYTNA--ARCCNTGGA-TNAC-NMGNCARAAYTTYWSNMGNTGYYTNGARYTNCA
     | ||  | || || |  |  |  | |   |||   ||  |  | || ||  || |   | || | ||| |
    GCCGACGACCTGGTCGACCGGATGACCACCGGACGCACCTCCGCCGACCTGGTCACCGAG-TTCGCGCTGCC
          590       600       610       620       630       640       650

           480       490       500       510       520       530
    R-TGYCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARG-CNAC-NGCNCCNAC
    | ||||  ||   ||| ||    || ||  |  |  || | |    || |  |||| | ||  |  |   |
    GCTGCCGTCC---CTGGTG--ATCTGCCTGCTGCTCGGCGTC-----CCCTACGAGGACCACGCGTTCTTCC
         660           670       680           690       700       710
```



```
540       550       560       570       580       590       600
  NGCNCCNCARCC-NCCNYTNYTNYTNYTNY-TNYTNYTNCC-NGTNGGNYTNYTNYTNYTNG---CNGCNGC
    |  | |||||     || ||    || |   |  || ||  |   | |     |  | |     |  ||
  AGGAGCGCAGCCGGGTCCTGCTCACCCTGCGGTCCACTCCCGAGGAAGTCCGGGCCGCCCAGGACGAGTTGC
     720       730       740       750       760       770       780

     610       620       630       640       650       660       670
  NTG-GTGYYTNCAYTGGC-ARM--GNACNMGNMGNMGNACNCCNMGNCCNGGNGARCARGTNCCNCCNGTNC
    | || |||   | |||  ||  | || |  ||  |  |  | | |   |  | ||    || || |  |
  TGGAGTACCTCGCCCGGCTCGCCCGGACCAAGCG-GGAGCGGCCGGACGACGCCATCATCAGCCGCCTGGTC
   790       800       810       820       830       840       850

      680       690       700    X
  CNWSNCCNCARGAYYTNYTNYTNGTNGARCAY
    | |  | || ||  |  |     || | |
  GCCCGCGGCGAGCTCGACGACACCCAGATCGCCACCATGGGA
  860       870       880       890


6. US-08-162-407-6 (1-705)
   Q27764       Adrenodoxin reductase gene (part 2).

 ID   Q27764 standard; DNA; 5089 BP.
 AC   Q27764;
 DT   19-NOV-1992  (first entry)
 DE   Adrenodoxin reductase gene (part 2).
 KW   Ad Red; unsequenced intron; cholesterol side chain cleavage system;
 KW   adrenodoxin; alternate splicing; ss.
 OS   Homo sapiens.
 FH   Key             Location/Qualifiers
 FT   intron          ..706
 FT   /*tag=  a
 FT   /number= 2
 FT   /label= "Intron continued from Q25050; see comments"
 FT   exon            707..799
 FT   /*tag=  b
 FT   /number= 3
 FT   intron          800..1012
 FT   /*tag=  c
 FT   /number= 4
 FT   exon            1013..1135
 FT   /*tag=  d
 FT   /number= 4
 FT   intron          1136..1336
 FT   /*tag=  e
 FT   /number= 4
 FT   exon            1337..1450
 FT   /*tag=  f
 FT   /number= 5
 FT   intron          1451..1757
 FT   /*tag=  g
 FT   /number= 5
 FT   exon            1758..1859
 FT   /*tag=  h
 FT   /number= 6
 FT   intron          1860..2623
 FT   /*tag=  i
 FT   /number= 6
 FT   exon            2624..2749
 FT   /*tag=  j
```

```
FT   /number= 7
FT   intron          2750..3005
FT   /*tag=  k
FT   /number= 7
FT   exon            3006..3090
FT   /*tag=  l
FT   /number= 8
FT   intron          3091..3218
FT   /*tag=  m
FT   /number= 8
FT   exon            3219..3418
FT   /*tag=  n
FT   /number= 9
FT   intron          3419..3498
FT   /*tag=  o
FT   /number= 9
FT   exon            3499..3670
FT   /*tag=  p
FT   /number= 10
FT   intron          3671..4319
FT   /*tag=  q
FT   /number= 10
FT   exon            4320..4490
FT   /*tag=  r
FT   /number= 11
FT   intron          4491..4617
FT   /*tag=  s
FT   /number= 11
FT   exon            4618..4748
FT   /*tag=  t
FT   /number= 12
FT   3'UTR           4749..5048
FT   /*tag=  u
FT   mat_peptide     ..4745
FT   /*tag=  v
FT   /note= "Peptide continued from Q25050; see comments"
FT   misc_signal     2641
FT   /*tag=  w
FT   /label= Alternative_splice_site
PN   WO9207865-A.
PD   14-MAY-1992.
PF   30-OCT-1991; U08081.
PR   31-OCT-1990; US-606876.
PA   (REGC ) UNIV CALIFORNIA.
PI   Lin D, Miller WL;
DR   WPI; 92-183618/22.
PT   Human adrenodoxin reductase gene and related polypeptide(s) - for
PT   diagnosis and treatment of diseases and genetic disorders causing
PT   high cholesterol levels
PS   Disclosure; Fig 2; 36pp; English.
CC   The sequence shows the second portion of the sequence of the human
CC   adrenodoxin reductase (Ad Red) gene.  A region of approx. 5kb of
CC   unsequenced DNA is represented by 50 n's.  The first portion of
CC   this sequence is given in Q25050.  There are a total of 12 exons and
CC   the whole sequence covers 6905 bp, excluding the unsequenced intron
CC   sequence.
CC   Ad Red is involved in the cholesterol side chain cleavage system
CC   and is loosely associated with the inner mitochondrial membrane.
CC   It recieves electrons from NADPH and then passes these electrons
CC   to adrenodoxin, which functions as an indiscriminate electron shuttle.
```



```
CC   Ad Red mRNA undergoes alternate splicing to produce two different
CC   forms, with the less abundant form containing 18 nuccleotides
CC   encoding 6 amino acids in the middle of the protein.  The
CC   alternative splicing site occurs at the begining of exon 7 (see
CC   feature table).
SQ   Sequence   5089 BP;    902 A;    1401 C;    1633 G;    1100 T;
SQ   53 Others;

Initial Score    =    118  Optimized Score  =    351  Significance =  6.23
Residue Identity =    25%  Matches          =    322  Mismatches   =   350
Gaps             =     72  Conservative Substitutions               =     0

            X       10        20        30        40        50        60
            ATGACNGTNYTNGCNCCNGCNTGGWSNCCNACNACNTAYYTNYTNYTNYTNYTNYT-NW
              |   |       | || || ||||      | ||  |||| |  |   | |  |  ||  |
   TCTTACTCTTCAGTAAGCCACATCCCCTGCTTGGATCATCCCCACTCACCTCCCACCAGT-TAACAGTTAAA
 2080      2090      2100      2110      2120      2130      2140

         70        80        90        100       110       120
   SNWSNGGNYTNWSNGGNACNCARGAYTGYWSNT---TYCARCAY-WSNCCNATNWSNWSNGAYTTYGCNGTN
   | ||    || ||      |  | |||| | |   ||||| || || |  |  || ||  ||| | |  |
   GTTCATCTTTGAGATCTTGGCTCGGTTGTCCCTTCCTCCAGGACTTCCCTGAATTGGTC-CACTCCCCCATC
2150      2160      2170      2180      2190      2200      2210

130       140       150       160       170       180       190
  AAR-ATNMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNW--SNAAYYTNCARGAY
    | | |    ||  ||   || ||| || | |  | |||   | |   | || |  | |||||   | ||
  GTGCCTCCTCCCATAGAC---CTGTCTACTTCTGTTTCACCGCAT-ATCATCGCCACGGCAATTT--TACAC
2220      2230      2240      2250      2260      2270      2280

 200       210       220       230       240       250       260
  GARGARYTNTGYGGNGGNYTNTGGMGNYTNGTNYT-NGCNCARMGNTGGATGGARMGNYTNAARACNG-TNG
       ||  ||  | |  || ||  |  |  | ||   | ||||    |||  ||| |   |||  |   |
  -----ACATTGATGAGTATT-TGATGAGTATTTCTCTCCCCAGC---AGATCACAAGCTCTGAGAGGGCCTG
      2290      2300      2310      2320      2330      2340

 270       280       290       300       310       320       330
  CNGGNWSNAARATGCARGGNYTNYTNGA--RMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTT
  |    ||      |     |  || ||  ||       | |  || | | || | |      | || | |
  CTCCTTGTCTCTTTTCCTCACCATT-GACCACCCCGCCCCTCCAAGCT-CTTTGCATGGGGGTGAGCACATG
2350      2360      2370      2380      2390      2400      2410

 340       350       360       370       380       390       400
  YCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMG-NYTNYT-NCARGARACNWSNGAR
    ||   ||  | || | |||  | |   |  ||          ||  |  || ||    |||| | |  | |
  GTAGAGGCCCACTACATACTT--GTTGAATGAAAGTGTGGGCCTGGTGTGCTATTAGTTGGAGCCATTTGCA
2420      2430      2440      2450      2460      2470      2480

 410       420       430       440       450       460       470
  CARYTNGTNGCNYTNAARCCNTGGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCAR---TGYCAR
   |||| |  ||  |     |     |   || ||   || |   || ||   ||| |||| | |     ||||
  GAGTTGGCAGCAATGCTCCATCTGGCCACCAGGGGGAGC-AGTGGCGCAATTT-GAGCCGCCAGCTCTCCAG
2490      2500      2510      2520      2530      2540      2550

  480       490       500       510       520       530       540
  CCNGAYWSNWSNACNYTNCCNCCNCCNTG--GWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCNC
   |  || |  | || |  ||        ||  ||| || |   | || | | | |  | |  |   | ||
  GCCCACGGTGGGACATGACC----AAATGGAGAGTCCCCTTTC-CTCG-GTC-AGTG-TCTGATGACTCCAA
2560      2570          2580      2590      2600          2610      2620
```

8. US-08-162-407-6 (1-705)
N60554 Sequence encoding bovine male specific sequence.

```
ID   N60554 standard; DNA; 5783 BP.
AC   N60554;
DT   24-MAY-1991  (first entry)
DE   Sequence encoding bovine male specific sequence.
KW   Plasmid pES8; PstI fragment; ss DNA; probe; sex determination;
OS   Bos sp.
PN   WO8607095-A.
PD   04-DEC-1986.
PF   30-MAY-1986; U01169.
PR   31-MAY-1985; US-739817.
PR   05-JUN-1986; US-871111.
PA   (SALK ) SALK INST FOR BIOL STUD.
PI   Ellis SB;
DR   WPI; 86-332096/50.
PT   Nucleic acid probes - for prenatal sexing of bovine species
PS   Disclosure; page 36-7; 75pp; English.
CC   The nucleic acid hybridises to a greater extent with total
CC   DNA of a breed of the genus Bos than with total DNA of females of the
CC   breed.  An embryo or foetus of the Bos species may be sexed by
CC   contacting cellular DNA with the labelled probe.  The probe provides
CC   rapid, 100% accurate sexing of bovine embryos using the small amt. of
CC   DNA obtd. from 4 embryonic cells.  Kits may be assembled.
SQ   Sequence   5783 BP;    745 A;    1627 C;    1029 G;    753 T;
SQ   1629 Others;
```

Initial Score = 117 Optimized Score = 366 Significance = 6.17
Residue Identity = 27% Matches = 341 Mismatches = 334
Gaps = 88 Conservative Substitutions = 0

```
            X       10          20        30        40        50
            ATGA-CNGTNYTNGCNCCNG--CNTGGWSNCCNACNACNTAYYTNYTNYTNYTNYTNYT
              ||| |    || || || | |  | ||||     |     |    || |     |  || |
   GGTGTCTGCACTGATCCAGGCTGGCTCCTGCCCTCGGTCAAGACCCTGGCAAACACTCCAGAGCCCACTCC-
  5030      5040      5050      5060      5070      5080      5090

  60        70        80            90       100       110       120
   NWSNWSNGGNYTNWSNGGNACNCARG--AYTG--YWSNTTYCARCA-YWSNCCNATNWSNWSNGAYTTYGCN
     | ||    |   | | |  | ||   |||  ||| ||| |||| ||| |   |  |  || | |
   -GGATCCCTGCACCGAAGCAGGCTGGCTCCTGCCTTCGTTCAAGCACCTGGCAAAGTA--ACCGCCGGGAGA
  5100      5110      5120      5130      5140      5150      5160

    130       140       150       160       170       180       190
   GTNAARATNMGNGARYTNWSNGAYTAYYTNYTNCARGAYT--AYCCNGTNACNGT-NGCNWSNAAYYTNCAR
    |  || | ||    ||  | | | ||||      |  ||  ||   ||  |  |  || || || ||   |
   CTCCAGCT-CGCTCCCTGCCCCGTCCACTTCGCTGTGTCCTGCATTTGGTCCCTCTCCGCGACAAAGCT-GTG
  5170      5180      5190      5200      5210      5220      5230

    200       210       220           230       240       250
   GAYGARGARYTNTGYGGNGGNYTNTGGMG-----NYTNGTNYTNGCNCARM-GNTGGA--TGGARMGNYTNA
     | | ||||   |   || |  || ||      || |  || |  | ||     ||    ||||  |  |
   ATCCCGCAGCTCACTCCCGG-CACTGAAGCCACCTCTGGGGCTTGGCCTGCTTCAAGACCCCGAGAAATACA
  5240      5250      5260      5270      5280      5290      5300

  260       270       280           290       300       310       320
   ARACNGTNGCNGGNWSNAARATGC-----ARGGNYT-NYTNGARMGNGTNAAYACNGARATNCAY-TTYGTN
```



```
     | || |  |     |||| ||     |||| ||  ||    ||    | | | | ||  ||| || |
   TCTCGGTCGGCGCACTGAAGAAGCAGTCTAGGGCCTGCCTTCTTCG----ATTCCCGCGACACACACTCCTG
5310      5320      5330      5340      5350          5360      5370

     330       340          350       360       370       380       390
A-CNAARTGYGCNTTYCARCC---NCCNCCNW--SNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTN
| | | |  ||| | |  |||     | |  |  | ||||  |   |||  ||| |   |   || |  ||
AGCCACAGCTGCGTGCGCGCCCTGAACTCGGATGGATGCTGCC--CTTGGGCAACC--CCCCCACCCTCCTC
5380      5390      5400      5410      5420      5430      5440

       400       410       420           430           440       450
YTNCARGARACNWSNGARCARYTNGT-NGCNYTNA-----ARCCNTG-GATNA--CNMGNCARAAYTTYWSN
|  | |  ||  ||  ||  |||  |  || || |       || || | | |  | ||   |  || |||
CCCCCACCAAACACACAAACACTCCTGAGCGTTCATCTCGGTCCCTGAGCTCAACCACGGTGG--TTCCTGC
  5450      5460      5470      5480      5490      5500      5510

     460       470       480       490          500       510       520
MGNTGYYTNGARYTNCARTGYCARCCNGAYWSNWSNACNY--TNCCNCCNCCNTGGWSNCCNMGNCCNYTNG
|  || |  |||   |   | ||| |  | || ||  | |    |   | || ||| | ||  | ||
CTCTGGC--GAAGGCCCCCGACAAAC--AATCCTGAGCCCAAACCTGGCCCCCTGGCCCCCTGGCCCCGAAC
  5520      5530      5540      5550      5560      5570      5580

      530       540       550       560       570       580       590
A--RGCNACNGCNCCNACNGCNCCNCARCCNCCNYTNYTNYTNYTNY-TNYTNYTNCCNGTNGGNYTNYTNY
|  ||  |   | ||   |  ||  | || ||    |  || || |   ||  | |   |    ||
AATGGTGAGTCCTCCCCTTGTTCC-AACCCCCCGGAATCACTGCTGTGCCCTCATCTCACGCGCTGAACTAA
    5590      5600      5610      5620      5630      5640      5650

      600       610       620       630       640           650
TNYTNGCNGCNGCNTGGTGYYTNCAYTGGCA-RMGNACNMGNMGNMGNAC-----NCCNMGNCCNGG----NG
  |  ||  | |  | |||||    ||  || |||  | || |  |  |     || || |  ||    |
ACCAAGC-ACGG--TCGTGCC----CTTCCAGAAGAGCCCGTCTAAATTCCTGAGCCCCGTCTCGGTCACTG
   5660          5670          5680      5690      5700      5710

  660       670       680       690       700    X
ARCARGTNCC--NCCNGTNCCNWSNCCNCAR-GAYYTNYTNYTNGTNGARCAY
|||  ||  ||   |  |  ||  | |    | || |  ||  |  | || | |
AAC-GGAACCAGTCGAGGTCCTGCCCTTGTGCGAGCGCCTAGTCCTTGACCGTAAGACATGAA
5720      5730      5740      5750      5760      5770      5780
```

9. US-08-162-407-6 (1-705)
Q06233 Bovine male-specific insert of lambda ES6.0 genome

```
ID   Q06233 standard; DNA; 4414 BP.
AC   Q06233;
DT   28-JAN-1991  (first entry)
DE   Bovine male-specific insert of lambda ES6.0 genome.
KW   Bovine male-specific insert; lambda ES6.0 embryo sexing;
KW   embryo transfer; ss.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   misc_RNA        466..469
FT   /*tag= a
FT   /note="best guess from ambiguous sequence info."
FT   misc_RNA        957..962
FT   /*tag= b
FT   /note="best guess from ambiguous sequence info."
FT   misc_RNA        1246..1252
FT   /*tag= c
```

```
 550       560       570       580           590       600       610
AR-CCNCCNYTNYTNYTNYTNYTNYTNCCNGTNGG----NYTNYTNYTNYTNGCNGCNGCNTGGTGYYT
 | || || ||  | |           |  |  |  ||      ||       | || |  || |  | | ||
GGCCCTCCTTTTGTGCCAGAGAACGGACATCACGAAGGCAGCCCTGGGTGTACT-GAGGCAGAGTCGAGTGA
    2630      2640      2650      2660      2670      2680      2690

       620       630       640           650       660           670
NCAY----TGGC-ARMGNACNMGNMGNMGNACNC--CNMG--NCCNGGNGAR--CARGTNC---CNCCNG-T
  ||     |||| || |  | || ||  |  | |  | ||    ||        |  |||| |      || |
AGACAGTGTGGCTAGTGGGC-CGGCGTGGCNCCCTGCAAGTGGCCTTCACCATTAAGGTGCTGGGGCCAGAG
   2700      2710      2720      2730      2740      2750      2760

       680       690       700    X
NCCNWSNCCNCARGAYYTNYTNYTNGTNGARCAY
    ||    |||| ||      |     | || |
TGGGAGGAGCCAGGGTTCCAGGTGGAGGGGGCCTCTCCTTAAGG
  2770      2780      2790      2800
```

7. US-08-162-407-6 (1-705)
   Q56917 Glutamic acid receptor.

```
 ID   Q56917 standard; DNA; 3717 BP.
 AC   Q56917;
 DT   08-AUG-1994  (first entry)
 DE   Glutamic acid receptor.
 KW   Glutamic acid receptor; synaptic signal translation; diagnosis;
 KW   brain disease; nerve signal; gene therapy; NMDA; cerebellum; ICR;
 KW   mouse; nerve cell necrosis; ds.
 OS   Mus musculus.
 FH   Key             Location/Qualifiers
 FT   CDS             1..3717
 FT   /*tag=  a
 FT   /product= glutamic_acid_receptor
 PN   J06014783-A.
 PD   25-JAN-1994.
 PF   30-JUN-1992; 173155.
 PR   30-JUN-1992; JP-173155.
 PA   (MITU ) MITSUBISHI KASEI CORP.
 DR   WPI; 94-061478/08.
 DR   N-PSDB; Q56916.
 PT   New glutamic acid receptor and gene - for use in analysis of
 PT   synaptic signal translation, and diagnosis of brain disease
 PS   Claim 1; Page 21-24; 35pp; Japanese.
 CC   The sequence codes for a glutamic acid receptor. The receptor is
 CC   useful for the analysis of nerve signal translation; within the
 CC   synapse, expression of synapse plasticity, nerve cell necrosis,
 CC   brain structure and brain disease. It can also be used in gene
 CC   therapy.
 SQ   Sequence   3717 BP;    675 A;    1199 C;    1120 G;    723 T;
```

```
Initial Score    =    117  Optimized Score  =    369  Significance =  6.17
Residue Identity =     29%  Matches          =    351  Mismatches   =   312
Gaps             =      96  Conservative Substitutions               =     0

             X       10          20        30        40        50        60
             ATGACNGTNYT-NGCNC-CNGCNTGGWSNCCNACNACNTAYYTNYTNYTNYTNYTNYTNYTN
                | | |  ||  || |   ||  ||||  | |   |  |  | |  |  |  |  | ||
    CCTTGGTGCTTGGGCAGGGCTGGGCGCAGGGCAGGGAGAAC-AGGCCGTGACCGTGGC-GGTGGTGTTTGGC
          50 X      60        70        80        90        100       110
```



```
          70        80          90        100       110       120
  WSNWSNGGNYTNWSNG-GNACNCARGAYTG--YWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTN
  || || || |  || |    | |||| | |  ||| | || || | |    |  || || |||||  | | |
  AGCTCTGGGC-CACTGCAGGCCCAGGCCCGGACTCGTCTC-ACCCCGCAGAACTTCCTG-GACTTGCCTCTG
        120       130       140       150       160       170       180

130       140       150       160       170       180       190       200
  AARATNMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGAR
  |||| |   |||  ||  || |||      ||| | ||||     ||    || || | ||| ||   | |
  GAGATCC---AGC-CACTCACCATCGGGGTCAACAATACC----AACCCCAGC-AGCATCCT-CA--CCCA-
            190       200       210       220       230

        210       220       230       240       250       260
  GARYTNTGYGGNGGNYT--NTGGMG-NYTNGTNYTNGCNCARMG--NTGGATGGARMGNYTNAARACNGTNG
  |||  |||||    ||   || ||  |  || |  |  || ||   | || ||| |    |  | || |  |
  -AAT-CTGTGGGCTCCTGGGTGCCGCCCGAGTCCACG-GCATCGTCTTTGA-GGACAACGTGGACACTGAGG
240       250       260       270       280       290       300

270       280       290       300       310       320       330
  CNGGNWSNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNA--ARTGYGCNTT
  | |   | || ||| ||  ||  |    |   |  | || |  | | | || | |  |  |||| |   |
  CCGTGGCTCAGCTGCTGGATTTCGT--CTCCTCT-CAGACCCACGTGCCCATCCTCAGCATCAGTG-GAGGT
  310       320       330       340       350       360       370

340               350       360       370       380       390       400
  YCARC---------CNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARAC
  || ||          | || || ||  |||  ||  ||    |   |  |||  ||  |  | | |  ||||
  TCTGCTGTGGTCCTCACCCCCAAGGAGCC--AGGCTCCGCCTTTC-TACAGCTG-GGCGTGTCCCTGGAG--
    380       390       400       410       420       430       440

        410       420       430       440       450       460
  NWSNGARCARYTNGTNGCNYTNAA-RCCNTGGA-TNACNMGNCARAAYTTYWSNMGNTGYYTNGA----RYT
   ||  ||| ||  ||     || || |   ||||   |  || | | |    || |||| |    ||
  CAG-CAGCTGCAGGTGCTGTTCAAGGTGCTGGAGGAATACGACTGGAGCGCGTTCGCTGTCATCACCAGCCT
         450       460       470       480       490       500       510

  470               480       490       500       510           520
  NCA---RTGYCA-RC---CNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCC-NMGNCC---NYTNGARGCN
   ||   | |||| ||   |  ||| ||  | |  | ||  |  ||| ||  |  ||     || | |||
  GCACCCGGGCCACGCGCTCTTCCTCGAGGGCGTGCGCGCCGTCGCCGACGCCAGCTACCTGAGCTGGCGGCT
       520       530       540       550       560       570       580

530       540       550           560       570       580       590
  ACNGCNCCNAC-NGCNCCNCARC----CNCCNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNYTNY
  | | |   |  | | |||    |  |  |     |       |      || | |   | || || |  |
  GCTGGACGTGCTCACGCTGGAGCTGGGCCCCGGTGGGCCGCGAGCGCGCACT--CAG--CGCTTACTGCGCC
     590       600       610       620       630           640       650

    600               610       620       630       640       650
  TNGCNGCNGC--------NTGGT-GYYTNCAYTGGCARMG-NACNMGNMGNM-GNACNCCNMGNCCNGG-NGA
    |  | ||          |||| ||||  ||||    ||  |  || || | |  | |  ||    ||  |
  AGGTCGACGCCCCGGTGCTGGTGGCCT--ACTGCTCCCGTGAAGAGGCGGAGGTGCTCTTCGCGGAGGCTGC
       660       670       680       690       700       710       720

  660       670       680       690       700    X
  RCARGTNCCNCCNGTNCCNWSNCCNCARGAYYTNYTNYTNGTNGARCAY
  |||||   |   |   |     | ||      || | |  || ||     ||
  ACAGG--CTGGC-TTGGTGGGACCCGGTCACGTGTGGTTAGTACCTAATCTGGCGCTGG
         730       740       750       760       770
```

```
FT   /note="best guess from ambiguous sequence info."
FT   misc_RNA        2912..3032
FT   /*tag=  d
FT   /note="only appears in about 50% of lambda ES6.0"
FT   misc_RNA        3080..3087
FT   /*tag=  e
FT   /note="best guess from ambiguous sequence info."
PN   US4960690-A.
PD   02-OCT-1990.
PF   05-JUN-1986; 871111.
PR   31-MAY-1985; US-739817.
PR   05-JUN-1986; US-871111.
PA   (SALK ) SALK INST FOR BIOL STUD.
PI   Ellis SB, Harpold MM;
DR   WPI; 90-319822/42.
PT   Bovine male-specific nucleic acid probes - used for sexing bovine
PT   embryos at the time of embryo transfer with high accuracy
PS   Example; Fig 1; 22pp; English.
CC   Lambda clone ES6.0 was isolated from a bovine liver-derived
CC   library. Filters were assayed for male-specific clones by comparing
CC   those that were pre-hybridised with sheared female bovine DNA and
CC   those that were not. Nine possible male-specific clones were
CC   identified for further analysis. In 2 of the 9 clones, bands
CC   corresponding to 4.0kb and 2.2kb, indicative of male-specific
CC   clones, were observed. ES6.0 was one of the clones. It was partially
CC   sequenced. Some bases could not be unambiguously identified. The long
CC   rows of 60 "N" residues within the sequence represent breaks in the
CC   6.2kb insert. The actual numbering of the blocks of sequence are
CC   420-716, 1089-1747, 2540-3560, 4000-6203.
CC   See also Q06234 and Q06235.
SQ   Sequence   4414 BP;    742 A;    1625 C;    1030 G;    756 T;
SQ   261 Others;

Initial Score     =    116  Optimized Score  =    367  Significance =  6.11
Residue Identity  =     28%  Matches         =    344  Mismatches   =   329
Gaps              =     92  Conservative Substitutions              =     0

            X        10          20        30        40        50
            ATGA-CNGTNYTNGCNCCNG--CNTGGWSNCCNACNACNTAYYTNYTNYTNYTNYTNYTNYT
              ||| |    || || || |  |  ||||      |       |    || |      |  || |
   GGTGTCTGCACTGATCCAGGCTGGCTCCTGCCCTCGGTCAAGACCCTGGCAAACACTCCAGAGCCCACTCC-
3660      3670      3680      3690      3700      3710      3720

  60        70        80            90        100         110           120
   NWSNWSNGGNYTNWSNGGNACNCARG--AYTG--YWSNTTYCARCA-YWSNCCNA-TNWS----NWSNGAYT
     | ||     |   | | |  | ||   |||  ||| ||| |||| ||| |  | | ||      | ||||
   -GGATCCCTGCACCGAAGCAGGCTGGCTCCTGCCTTCGTTCAAGCACCTGGCAAAGTAACCGCCGGGAGACT
3730      3740      3750      3760      3770      3780      3790      3800

         130       140       150       160           170       180
   TY-GCNGTNAARATNMGNGARYTNWSNGAYTAYYTNYTNCA---RGAYTAYC-CNGTNACNGTNGCNWSNAA
    | ||        | |  |  |  |  | || | | || ||   || |  || | |  ||    ||     |
   CCAGCTCGCTCCCTGCCCGTCCACTTCG-CTGTGTCCTGCATTTGGTCCCTCTCCGGGAC-AAAGCTGTGAT
        3810      3820      3830      3840      3850      3860      3870

  190         200       210       220       230       240       250
   YYTNCAR--GAYGARGARYTNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTN
   ||  |||   ||     ||  ||  |    || ||| | || |  |  ||   |||       ||||  |
   CCCGCAGCTCACTCCCGGCACTGAAGCCACCTCTGGGGCTTGG--CCTGCTTCAAG--ACCCCGAGAAATAC
        3880      3890      3900      3910      3920      3930
```



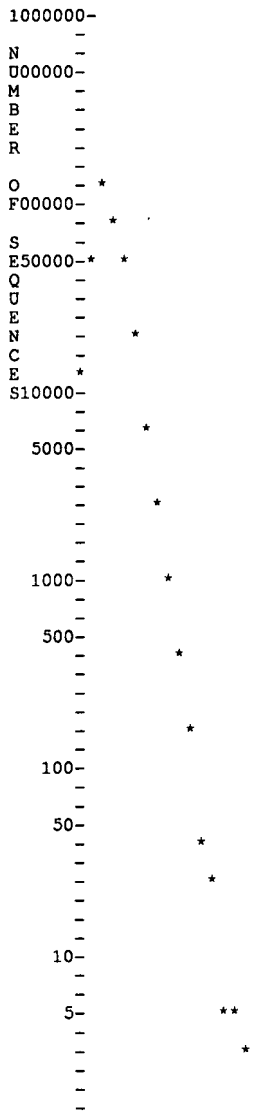
```
 260       270       280           290       300       310       320
  AARACNGTNGCNGGNWSNAARATGC-----ARGGNYT-NYTNGARMGNGTNAAYACNGARATNCAY-TTYGT
  |   | || |  |      |||| ||      |||| ||  ||     ||     | | | | ||  |||  || |
  ATCTCGGTCGGCGCACTGAAGAAGCAGTCTAGGGCCTGCCTTCTTCG----ATTCCCGCGACACACACTCCT
3940      3950      3960      3970      3980           3990      4000

       330       340           350       360       370       380
  NA-CNAARTGYGCNTTYCARCC---NCCNCCNW--SNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYT
   | | |  ||| | |  |||      | |  |  | ||||  |   |||  ||| |   |   || |  ||
  GAGCCACAGCTGCGTGCGCGCCCTGAACTCGGATGGATGCTGCC--CTTGGGCAACC--CCCCCACCCTCCT
  4010      4020      4030      4040      4050          4060      4070

390       400       410       420       430       440       450
  NYTNCARGARACNWSNGARCARYTNGT-NGCNYTNAARCCNTGGATNACNMG-NCAR-AAYTTYWSNMGNTG
   |  | |  ||  ||  ||  ||| |  || ||   | |  ||        || |||  ||  | |  ||
  CCCCCCACCAAACACACAAACACTCCTGAGCGTT--CATCTCGGTCCCTGAGCTCAACCACGGTGGTTCCTG
    4080      4090      4100          4110      4120      4130      4140

 460       470       480       490       500       510       520
  YYTNGARYTNCARTGYCARCCNGAYWSNWSNA-CNYTNCCNCCNCCNTGGWSNCCNMGNCCNYT-----NGA
  |||   ||  ||| |||  || ||||  || | | |   | |   | ||| | ||  | || ||      ||
  CCTCTGGC--CAAGGCC--CCCGACA-AACAATCCTGAGCCCAAACCTGGCCCCCTGGCCCCCTGGCCCCGA
    4150          4160      4170      4180      4190      4200      4210

         530       540       550       560       570       580       590
  R-----GCNACNGCNCCNACNGCNCCNCARCCNCCNYTNYTNYTNYTNY-TNYTNYTNCCNGTNGGNYTNYT
  |     |  |   | ||     |  ||  | || ||     |  || || |   ||  |  |     |  ||
  ACAATGGTGAGTCCTCCCCTTGTTCC-AACCCCCCGGAATCACTGCTGTGCCCTCATCTCACGCGCTGAACT
       4220      4230      4240      4250      4260      4270      4280

         600       610       620       630       640           650
  NYTNYTNGCNGCNGCNTGGTGYYTNCAYTGGCA-RMGNACNMGNMGNMGNAC-----NCCNMGNCCNGG----
      |  ||  | |  | |||||    ||  || ||| | || |  |  |     || || |  ||
  AAACCAAGC-ACGG--TCGTGCC----CTTCCAGAAGAGCCCGTCTAAATTCCTGAGCCCCGTCTCGGTCAC
      4290          4300          4310      4320      4330      4340

      660       670       680       690       700    X
  NGARCARGTNCC--NCCNGTNCCNWSNCCNCAR-GAYYTNYTNYTNGTNGARCAY
   |||| ||  ||   |  |  || | |   | || |  ||  |  | || | |
  TGAAC-GGAACCAGTCGAGGTCCTGCCCTTGTGCGAGCGCCTAGTCCTTGACCGTAAGACATGAA
 4350      4360      4370      4380      4390      4400      4410
```

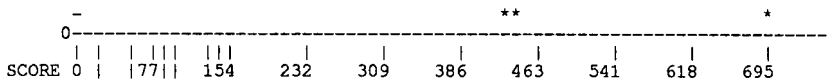
```
10. US-08-162-407-6 (1-705)
    Q02830       cDNA to HIV-2 RNA.

ID   Q02830 standard; cDNA; 9360 BP.
AC   Q02830;
DT   29-MAY-1989  (first entry)
DE   cDNA to HIV-2 RNA.
KW   HIV; AIDS; Vaccine; pUC-HIV-2(GH-1).
OS   HIV-2.
FH   Key             Location/Qualifiers
FT   CDS             544..2109
FT   /*tag=  a
FT   /label=Gag reading frame
FT   CDS             1830..4934
FT   /*tag=  b
FT   /label=Pol gene reading frame.
```

```
FT   CDS             4867..5511
FT   /*tag=  c
FT   /label=Q gene reading frame
FT   CDS             5342..5677
FT   /*tag=  d
FT   /label=X gene reading frame
FT   CDS             5680..5994
FT   /*tag=  e
FT   /label=R gene reading frame
FT   CDS             6148..8703
FT   /*tag=  f
FT   /label=Env gene reading frame.
FT   CDS             8540..9304
FT   /*tag=  g
FT   /label=F gene reading frame.
PN   J01289486-A.
PD   21-NOV-1989.
PF   16-MAY-1988; 119024.
PR   16-MAY-1988; JP-119024.
PA   Toa Nenryo Kogyo KK,  Fuji Kebio KK.
DR   WPI; 90-005177/01.
DR   P-PSDB; R04024-30.
PT   DNA indicating complement to RNA gene -
PT   of Human Immunodeficiency Virus type 2 used for new vaccine or
PT   diagnostic for AIDS virus.
PS   Claim 2; Fig.4; 12pp; Japanese.
CC   cDNA to novel HIV-2 (GH-1) has been integrated into plasmid
CC   pUC HIV-2(GH-1). Useful for diagnosis and vaccination against the virrus.
CC   Carries 7 overlapping genes in varrious reading frames, includding Gag,
CC   Pol and Env.
SQ   Sequence   9360 BP;   3211 A;   1944 C;   2317 G;   1888 T;

Initial Score     =    116  Optimized Score  =    348  Significance =  6.11
Residue Identity  =    26%  Matches          =    318  Mismatches   =   355
Gaps              =     67  Conservative Substitutions              =     0

            X       10          20          30          40          50
            ATGACNGTNY--TNGCNCCNGCNTGG--WSNCCNACNACNT-AYYTNYTNYTNYTNYTNYTN
            ||||   | |  | ||    |   ||  ||    |   | | ||    |  | || |  ||
   CAATACATGGATGATATTTTAATAGCTAGTGACAGGACAGGTTTAGAGCATGACAAAGTAGTCCTGCAGCT-
       2930 X    2940        2950        2960        2970        2980        2990

    60          70          80          90          100         110         120
   YTNWSNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAY----WSNCCNATNWSNWSNGAYTTYGC
      || ||     || |  ||  |  ||| || |||   || | ||    ||  | |  |  ||   || |||
   AAAAGAAC--TTCTGAATGG--CCTAGGGTT-CTCTACCCCAGATGAGAAGTTCCA--AAAAGACCCTCCGC
       3000        3010        3020        3030        3040        3050

     130         140         150         160         170         180         190
   NGTNAA--RATNMGNGARYTNWSNGAYTAYYTNYTNCARGAY-TAYCCNGTNACNGTNGCNWSNAAYYTNCA
     | ||  |||    | ||| ||  |||||   |  | | ||  |  | |   | |     || | |   |
   -TTCAATGGAT----GGGCTATG-AACTATGGCCAACTAAATGGAAGCTG---CAGAAATTACAACTGCCCC
   3060         3070        3080        3090        3100         3110        3120

      200         210         220         230         240         250         260
   RGAYGARGARYTNTGYGGNGGNYTNTGGMGNYTNG-TNYTNG-CNCARMGNTGGATGGARMGNYTNAARACN
   |||  ||||| | || |   | |   | |  |  |   || |    ||    |  || | ||| |  ||| |
   AGA--AAGAAATATG-GACAGTCAACGACATCCAGAAGCTGGTGGGAGTCTTAAATTGGGCGGCACAAATCT
           3130        3140        3150        3160        3170        3180
```



```
       270         280         290         300         310         320         330
   GTNGCNGGNWSNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARAT-NCAYTTYGTNACNAARTGYGC
     |  | || |  ||||   ||   || |    ||   | ||     ||||  ||||      |  ||||||
   AT-CCAGGAATAAAAACCAAACATTTGTGTAGGCTAATCAAAGGAAAAATGACACTCACAGAAGAAGTGCAG
3190        3200        3210        3220        3230        3240        3250        3260

      340         350         360         370         380         390         400
   NTTYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAA--YATNWSNMGNYTNYTNCARGARACNWSN
       |||       |  ||  | |  || |      |||| ||  ||| |  || |     ||||||   |
   TGGACAGAGTTAGC--AGAAG-CAGAGCTAGAAGAAAACAAAATTATCTTAAGCCAGGAACAAGAAGGATAC
         3270        3280        3290        3300        3310        3320

     410         420         430         440         450         460         470
   GARCARYTNGTNGCNYTNAARCCNTGGA--TNACNMGNC-ARAAYTTYW-SNMGNTGYYTNGARYTNCARTG
    |  | |  |  |      |||   | ||    || |  | ||||  ||| | |  |   | ||  | |||
   TATTACCAAGAAG--AAAAAGAATTAGAGGCAACAATCCAAAAAAAATCAAGACAATCAGTGGACATACAAAA
3330        3340        3350        3360        3370        3380        3390

       480         490         500         510         520         530         540
   YCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGG--WSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCN
   |  ||| |  || || |  |  ||        || |      |     |   ||    |  || |  |
   TACACCAGGAAG-AGAAAAT--CCTAAAAGTAGGGAAATATGCAAAGATAAAAAATACCCATACCAATGGGG
3400        3410        3420        3430        3440        3450        3460

         550         560         570         580         590         600         610
   CCNCA--RCCNCCNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNYTNYTNGCNGCNG-CNTGGTGY
     |  |  ||   | |  |  | | |      |      |  |  || | |  || | |  | |
   TCAGATTACTAGCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTAGT-CATCTGGGGACGGATACCAAAA
 3470        3480        3490        3500        3510        3520        3530

       620         630         640         650         660         670         680
   YTNCAYTGGCARMGNACNMGNMGNMGNACNCCNMGNCCNGG-NGARCARGTNC-CNCCNGTNCCNW-SNCCN
   || ||||  | |       || || ||  |    ||    ||  | | || | |   | ||  | | |  |
   TTCCACTTACCGGTGGAGAG-AG-AGACCTGGGAGCAGTGGTGGGATAACTACTGGCAAGTGACATGGATCC
3540        3550        3560        3570        3580        3590        3600

              690         700    X
   CA-----RGAYYT-NYTNYTNGTNGARCAY
   ||     |||||| |  ||      | |||
   CAGAGTGGGACTTCGTATCTACCCCACCACTAGTCAGGTT
3610        3620        3630        3640
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file transl.res made by root on Thu 6 Apr 95 9:53:08-PDT.


Query sequence being compared:US-08-162-407-6 (1-705)
Number of sequences searched:                313646
Number of scores above cutoff:                 4021

     Results of the initial comparison of US-08-162-407-6 (1-705) with:
   Data bank : EMBL-NEW 1, all entries
   Data bank : GenBank 86, all entries
   Data bank : GenBank-NEW 1, all entries
   Data bank : UEMBL 41_86, all entries
```





PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	33	33	13.99

Times:	CPU	Total Elapsed
	00:39:05.03	00:39:16.00

Number of residues: 288682119
Number of sequences searched: 313646
Number of scores above cutoff: 4021

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 47 standard deviations above mean ****					
1. HSU03858	Human flt3 ligand mRNA, compl	1080	695	705	47.31	0
2. HSU04806	Human FLT3/FLK2 ligand mRNA,	859	693	704	47.16	0
	**** 29 standard deviations above mean ****					
3. MUSLIGAND	Mouse flt3 ligand mRNA, compl	829	442	586	29.23	0
	**** 28 standard deviations above mean ****					
4. MMU04807	Mus musculus FLT3/FLK2 ligand	1152	427	578	28.15	0
	**** 10 standard deviations above mean ****					
5. SCMIG1	S.cerevisiae MIG1 gene for a	2622	173	368	10.00	0
	**** 9 standard deviations above mean ****					
6. TAGLUIN1	Wheat Glu-Aly sublocus with W	794	164	296	9.36	0
7. TAGLU1AG	Wheat gene for HMW-glutenin s	2915	164	362	9.36	0
8. HSCOL16A	Homo sapiens alpha-1 type XVI	5387	160	306	9.08	0
	**** 8 standard deviations above mean ****					
9. SCBYP13CS	S.cerevisiae gene for bypl-3	4100	156	368	8.79	0
10. WHTGLIABF	Wheat (T.aestivum) alpha-/bet	950	153	227	8.58	0
11. DUKMTTGRA	Anas platyrhyncos mitochondri	2046	153	364	8.58	0
12. SQDSKLC	Loligo pealii kinesin light c	2043	152	371	8.50	0
13. A10377	Artificial sequence for prepr	1368	147	368	8.15	0
14. XELSPAA	X.laevis spasmolysin mRNA.	1378	147	368	8.15	0

```
                ****  7 standard deviations above mean ****
  15. ADRE3FIBES    Adenovirus 11 E3 region encod   2481   143   324   7.86   0
  16. MMOCT2710     M.musculus Oct-2 gene.          1246   140   382   7.65   0
  17. HUMDSP        Human duhydrolipoamide succin   3012   139   397   7.57   0
  18. HUMCOL16A     Homo sapiens alpha-1 type XVI   5387   139   306   7.57   0
  19. HUMCOL16A     Homo sapiens alpha-1 type XVI   5387   139   306   7.57   0
  20. SCYKR007W     S.cerevisiae chromosome XI re    950   138   364   7.50   0


1. US-08-162-407-6 (1-705)
   HSU03858     Human flt3 ligand mRNA, complete cds.

 LOCUS       HSU03858     1080 bp    mRNA            PRI       19-JUL-1994
 DEFINITION  Human flt3 ligand mRNA, complete cds.
 ACCESSION   U03858
 KEYWORDS    .
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 1080)
   AUTHORS   Lyman,S.D., James,L., Johnson,L., Brasel,K., de Vries,P.,
             Escobar,S.S., Downey,H., Splett,R.R., Beckmann,M.P. and
             McKenna,H.J.
   TITLE     Cloning of the human homologue of the murine flt3 ligand: a growth
             factor for early hematopoietic progenitor cells
   JOURNAL   Blood 83, 2795-2801 (1994)
   STANDARD  full automatic
 REFERENCE   2  (bases 1 to 1080)
   AUTHORS   Lyman,S.D.
   TITLE     Direct Submission
   JOURNAL   Submitted (30-NOV-1993) Stewart D. Lyman, Immunex Research and
             Development Corporation, 51, University St., Seattle, WA 98101, USA
   STANDARD  full automatic
 COMMENT     NCBI gi: 494978
 FEATURES             Location/Qualifiers
     source           1..1080
                      /clone_lib="expression, cDNA"
                      /organism="Homo sapiens"
                      /cell_line="clone 22 (T cell)"
     5'UTR            1..83
     CDS              84..791
                      /standard_name="FMS-like tyrosine kinase-3 ligand"
                      /note="ligand for the flt3/flk-2 tyrosine kinase receptor;
                       NCBI gi: 494979"
                      /codon_start=1
                      /function="stimulates proliferation of early hematopoietic
                      cells"
                      /product="flt3 ligand"
                      /translation="MTVLAPAWSPTTYLLLLLLLSSGLSGTQDCSFQHSPISSDFAVK
                      IRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTVAGSKMQGLLERVN
                      TEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQP
                      DSSTLPPPWSPRPLEATAPTAPQPPLLLLLLLPVGLLLLAAAWCLHWQRTRRRTPRPG
                      EQVPPVPSPQDLLLVEH"
     sig_peptide      84..161
     misc_feature     162..629
                      /note="extracellular domain"
     misc_feature     630..698
                      /note="transmembrane domain"
     misc_feature     699..788
```



```
                      /note="cytoplasmic domain"
     3'UTR            792..1080
     misc_feature     1015..1080
                      /note="ATTTA mRNA instability motif"
     polyA_signal     1059..1064
     polyA_site       1080
                      /note="32 A residues"
 BASE COUNT      204 a    384 c    290 g    202 t
 ORIGIN

Initial Score    =    695  Optimized Score  =    705  Significance = 47.31
Residue Identity =    54%  Matches          =    545  Mismatches   =   160
Gaps             =      0  Conservative Substitutions               =     0

            X       10        20        30        40        50        60
             ATGACNGTNYTNGCNCCNGCNTGGWSNCCNACNACNTAYYTNYTNYTNYTNYTNYTNWS
             ||||| || || || || || ||||| || || || ||||| || || || || || || ||
   CCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCAACAACCTATCTCCTCCTGCTGCTGCTGAG
        80   X    90        100       110       120       130       140

         70        80        90        100       110       120       130
   NWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARAT
    || || || || || || ||||||||||| |||||||||||| || || || || |||||||| || |||||
   CTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAAAAT
     150       160       170       180       190       200       210

      140       150       160       170       180       190       200
   NMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYT
    || |||||| || |||||||| || |||||||||| || || || || || |||||| ||||||||||||
   CCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCT
   220       230       240       250       260       270       280

    210       220       230       240       250       260       270
   NTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNAA
    |||||| || || ||||| || || || || ||||| |||||||||||| || ||||| || || || || ||
   CTGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAA
290       300       310       320       330       340       350       360

280       290       300       310       320       330       340       350
 RATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCNCC
 |||||||| || || ||||| || ||||| ||||| |||||||| || |||||||| |||||||| || ||
 GATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCC
       370       380       390       400       410       420       430

        360       370       380       390       400       410       420
 NWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGCNYT
  || ||||| || ||||| ||||| ||||| || || || || |||||||| || |||||||| || || ||
 CAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCT
     440       450       460       470       480       490       500

      430       440       450       460       470       480       490
 NAARCCNTGGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWSNWSNACNYT
  ||||| ||||| || || |||||||||| || ||||| ||||| |||||||||| ||||| || || ||
 GAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTCAACCCT
   510       520       530       540       550       560       570

     500       510       520       530       540       550       560
 NCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCNCARCCNCCNYTNYTNYTNYT
  || || || ||||| || || || || ||||| || || || || || ||||| || || || || || ||
 GCCACCCCCATGGAGTCCCCGGCCCCTGGAGGCCACAGCCCCGACAGCCCCGCAGCCCCCTCTGCTCCTCCT
```

```
    580       590       600       610       620       630       640

     570       580       590       600       610       620       630
    NYTNYTNYTNCCNGTNGGNYTNYTNYTNYTNGCNGCNGCNTGGTGYYTNCAYTGGCARMGNACNMGNMGNMG
     || || || || || || || || || || || || || ||||||||| |||||||||| || || || ||
    ACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCGGAG
  650       660       670       680       690       700       710       720

   640       650       660       670       680       690       700     X
    NACNCCNMGNCCNGGNGARCARGTNCCNCCNGTNCCNWSNCCNCARGAYYTNYTNYTNGTNGARCAY
     || || || || || |||||||| || || || || || || |||||||.| || || || ||||||
    GACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCACTGACC
         730       740       750       760       770       780       790

    TGGCC
```

2. US-08-162-407-6 (1-705)
   HSU04806 Human FLT3/FLK2 ligand mRNA, complete cds.

```
LOCUS       HSU04806      859 bp    mRNA            PRI       11-MAY-1994
DEFINITION  Human FLT3/FLK2 ligand mRNA, complete cds.
ACCESSION   U04806
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
            Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
            Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
            Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
            Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
  TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
            haematopoietic stem cells and is encoded by variant RNAs
  JOURNAL   Nature 368, 643-648 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 859)
  AUTHORS   Culpepper,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
            Institute, 901 California Avenue, Palo Alto, CA 94304, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 483844
FEATURES             Location/Qualifiers
     source          1..859
                     /clone="human Flt3/Flk2 ligand S86"
                     /clone_lib="SV48 cDNA"
                     /organism="Homo sapiens"
                     /cell_line="SV48 thymic stroma"
     CDS             93..800
                     /note="NCBI gi: 483845"
                     /codon_start=1
                     /product="FLT3/FLK2 ligand"
                     /translation="MTVLAPAWSPTTYLLLLLLLSSGLSGTQDCSFQHSPISSDFAVK
                     IRELSDYLLQDYPVTVASNLQDEELCGALWRLVLAQRWMERLKTVAGSKMQGLLERVN
                     TEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQP
                     DSSTLPPPWSPRPLEATAPTAPQPPLLLLLLLPVGLLLLAAAWCLHWQRTRRRTPRPG
```



```
                    EQVPPVPSPQDLLLVEH"
 BASE COUNT      147 a    318 c    246 g    148 t
 ORIGIN

Initial Score    =    693  Optimized Score  =    704  Significance = 47.16
Residue Identity =    54%  Matches          =    544  Mismatches   =   161
Gaps             =      0  Conservative Substitutions               =     0

            X       10        20        30        40        50        60
           ATGACNGTNYTNGCNCCNGCNTGGWSNCCNACNACNTAYYTNYTNYTNYTNYTNYTNWS
            ||||| || || || || || ||||| || || || ||||| || || || || || || ||
   CCCGGCCGAAATGACAGTGCTGGCGCCAGCCTGGAGCCCAACAACCTATCTCCTCCTGCTGCTGCTGCTGAG
         90  X    100       110       120       130       140       150

        70        80        90       100       110       120       130
   NWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARAT
    || || || || || || |||||||||| |||||||||| || || || || |||||||| || |||||
   CTCGGGACTCAGTGGGACCCAGGACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAAAAT
      160       170       180       190       200       210       220

      140       150       160       170       180       190       200
   NMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYT
    || ||||| || |||||||| || |||||||||| || || || || || ||||| ||||||||||||||
   CCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCT
    230       240       250       260       270       280       290

    210       220       230       240       250       260       270
   NTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNAA
    ||||| |  || ||||| || || || || ||||| |||||||||| || ||||| || || || || ||
   CTGCGGGGCGCTCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAA
300       310       320       330       340       350       360       370

280       290       300       310       320       330       340       350
 RATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCNCC
 ||||||||| || || ||||| || ||||| ||||| |||||||| || |||||||| |||||||| || ||
 GATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCC
       380       390       400       410       420       430       440

       360       370       380       390       400       410       420
 NWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGCNYT
  || ||||| || ||||| ||||| ||||| || || || || |||||||| || |||||||| || || ||
 CAGCTGTCTTCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCT
      450       460       470       480       490       500       510

     430       440       450       460       470       480       490
 NAARCCNTGGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWSNWSNACNYT
  ||||| ||||| || || |||||||||| || ||||| ||||| |||||||||| ||||| || || ||
 GAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTCAACCCT
    520       530       540       550       560       570       580

    500       510       520       530       540       550       560
 NCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCNCARCCNCCNYTNYTNYTNYT
  || || || ||||| || || || || ||||| || || || || || || || ||||| || || || || ||
 GCCACCCCCATGGAGTCCCCGGCCCCTGGAGGCCACAGCCCCGACAGCCCCGCAGCCCCCTCTGCTCCTCCT
  590       600       610       620       630       640       650

 570       580       590       600       610       620       630
 NYTNYTNYTNCCNGTNGGNYTNYTNYTNYTNGCNGCNGCNTGGTGYYTNCAYTGGCARMGNACNMGNMGNMG
  || || || || || || || || || || || || || |||||||| |||||||||| || || || ||
 ACTGCTGCTGCCCGTGGGCCTCCTGCTGCTGGCCGCTGCCTGGTGCCTGCACTGGCAGAGGACGCGGCGGAG
```

```
    660       670       680       690       700       710       720       730

    640       650       660       670       680       690       700     X
     NACNCCNMGNCCNGGNGARCARGTNCCNCCNGTNCCNWSNCCNCARGAYYTNYTNYTNGTNGARCAY
      || || || || || ||||||||| || || || || || || ||||||||| || || || ||||||
     GACACCCCGCCCTGGGGAGCAGGTGCCCCCCGTCCCCAGTCCCCAGGACCTGCTGCTTGTGGAGCACTGACC
           740       750       760       770       780       790       800

     TGGCC
```

3. US-08-162-407-6 (1-705)
   MUSLIGAND Mouse flt3 ligand mRNA, complete cds.

```
 LOCUS       MUSLIGAND     829 bp ss-mRNA            ROD       04-JAN-1994
 DEFINITION  Mouse flt3 ligand mRNA, complete cds.
 ACCESSION   L23636
 KEYWORDS    ligand.
 SOURCE      Mus musculus (strain SJL/J) cDNA to mRNA.
   ORGANISM  Mus musculus
             Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
             Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
 REFERENCE   1  (bases 1 to 829)
   AUTHORS   Lyman,S.D., James,L., VandenBos,T., deVries,P., Brasel,K.,
             Gliniak,B., Hollingsworth,L., Picha,K.S., McKenna,H.J.,
             Splett,R.R., Fletcher,F.A., Maraskovsky,E., Farrah,T.,
             Foxworthe,D., Williams,D.E. and Beckmann,M.P.
   TITLE     Molecular cloning of a ligand for the flt3/flk-2 tyrosine kinase
             receptor -- a proliferative factor for primitive hematopoietic
             cells
   JOURNAL   Cell 75, 1157-1167 (1993)
   STANDARD  full automatic
 COMMENT     NCBI gi: 439441
 FEATURES             Location/Qualifiers
      source          1..829
                      /organism="Mus musculus"
                      /strain="SJL/J"
                      /sequenced_mol="cDNA to mRNA"
      sig_peptide     32..109
                      /codon_start=1
      CDS             32..727
                      /note="NCBI gi: 439442"
                      /codon_start=1
                      /product="flt3 ligand"
                      /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                      KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                      NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLALKPCIGKACQNFSRCLEVQ
                      CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLPLTLVLLAAAWGLRWQRARRR
                      GELHPGVPLPSHP"
      mat_peptide     110..724
                      /codon_start=1
 BASE COUNT      163 a    265 c    221 g    180 t
 ORIGIN

Initial Score    =    442  Optimized Score  =    586  Significance = 29.23
Residue Identity =    44%  Matches          =    465  Mismatches   =   239
Gaps             =     20  Conservative Substitutions               =     0

             X       10        20        30        40        50
```



```
                ATGACNGTNYTNGCNCCNGCNTGGWSNCCNACNACNTAY---YTNYTNYTNYTNYTNYT
                ||||| || || || || || ||||| || |   | | |   || || || || || || ||
   CCCGGCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCT
          30 X      40        50        60        70        80        90

  60        70        80        90       100       110       120       130
   NWSNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAA
    ||  |  | ||  | || || |  |||||||  |||   ||||| || || || ||  |||||   || ||
   GAGTCCTTGCCTGCGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAA
       100       110       120       130       140       150       160

        140       150       160       170       180       190       200
   RATNMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGA
   | | || ||||| || ||| |||| ||  |||||||||| || || || ||    ||||| |||||||||| |
   GTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAGAA
     170       180       190       200       210       220       230

       210       220       230       240       250       260       270
   RYTNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWS
   ||  |||   |  || ||||| ||  | || || ||||| ||||| ||||  || ||||| || || || ||
   GCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTC
   240       250       260       270       280       290       300

     280       290       300       310       320       330       340
   NAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCC
    |||||||||   || || |||   || ||||| ||||| |||||||| ||   |||| | ||||||||| |
    TAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCT
310       320       330       340       350       360       370       380

     350       360       370       380       390       400       410
   NCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGC
    ||    ||||| || ||||| ||||| ||||| || |  || ||  |||| || ||   ||||||  | ||
    ACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCTGCTTGC
          390       400       410       420       430       440       450

420       430               440       450       460       470       480
   NYTNAARCCNT------GGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWS
    || ||||| |       |||   |  | ||||||||||| || ||||| ||| | ||||||||||| |||||
   TCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTC
       460       470       480       490       500       510       520

     490       500       510       520       530               540       550
   NWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNAC--NGCNCCNA---CNGCNCCNCA-RCC
    || || ||·|  || ||  |||| || |   | || ||||| ||   || || |   |  |  | || ||
   CTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCA
     530       540       550       560       570       580       590

          560       570       580       590       600       610       620
   NCCNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNYTNYTNGCNGCNGCNTGGTGYYTNCAYTGGCA
    |  || || || || || || || || ||  |    ||  | || || || || ||| |||| | ||||||
   GCTGTTGCTCCTGCTGCTGCTGCTGCCTCTCACACTGGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTGGCA
   600       610       620       630       640       650       660

       630           640       650       660       670       680       690
   RMGNACNMGNMGNM---GNACNCCNMGNCCNGGNGARCARGT-NCCNCCNGTNCCNWSNCCNCARGAYYTNY
   |||  | || || |   |    |  | || || | ||   |  || ||  | || || |     | || ||| |
    AAGGGCAAGAAGGAGGGGGGAGCTCCACCCTGGGGTGCCCCTCCCCTCCCATCCCTAGGATTCGAGCCTTGT
670       680       690       700       710       720       730       740

          700    X
```

```
    TNYTNGTNGARCAY
       | || | || |
    GCATCGTTG-ACTCAGCCAGGGTC
          750    X   760
```

4. US-08-162-407-6 (1-705)
   MMU04807 Mus musculus FLT3/FLK2 ligand mRNA, complete cds.

```
LOCUS       MMU04807     1152 bp    mRNA            ROD       11-MAY-1994
DEFINITION  Mus musculus FLT3/FLK2 ligand mRNA, complete cds.
ACCESSION   U04807
KEYWORDS    .
SOURCE      mouse.
  ORGANISM  Mus musculus
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Rodentia; Myomorpha; Muridae; Mus.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
            Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
            Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
            Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
            Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
  TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
            haematopoietic stem cells and is encoded by variant RNAs
  JOURNAL   Nature 368, 643-648 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1152)
  AUTHORS   Culpepper,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
            Institute, 901 California Avenue, Palo Alto, CA 94304, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 483846
FEATURES             Location/Qualifiers
     source          1..1152
                     /clone="mouse Flt3/Flk2 ligand T110"
                     /clone_lib="TA4 cDNA"
                     /organism="Mus musculus"
                     /cell_line="TA4 stroma"
     CDS             256..954
                     /note="NCBI gi: 483847"
                     /codon_start=1
                     /product="FLT3/FLK2 ligand"
                     /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                     KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                     NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLGLKPCIGKACQNFSRCLEVQ
                     CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLLPLTLVLLAAAWGLRWQRARR
                     RGELHPGVPLPSHP"
BASE COUNT      238 a    346 c    324 g    244 t
ORIGIN
```

```
Initial Score    =    427  Optimized Score  =    578  Significance = 28.15
Residue Identity =    43%  Matches          =    458  Mismatches   =   244
Gaps             =     21  Conservative Substitutions               =     0

             X       10        20        30           40        50
             ATGACNGTNYTNGCNCCNGCNTGGWSNCCNACNACNTAY---YTNYTNYTNYTNYTNYT
             ||||| || || || || || || ||||| || |   | | |   || || || || || || ||
   CCCGGCAGAGATGACAGTGCTGGCGCCAGCCTGGAGCCCAAATTCCTCCCTGTTGCTGCTGTTGCTGCTGCT
```



```
     250     X 260       270       280       290       300       310

 60        70        80        90       100       110       120       130
  NWSNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAA
   ||  |  | ||  | || || |  ||||||| |||   ||||| || || || ||  |||||   || ||
  GAGTCCTTGCCTGCGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAA
  320       330       340       350       360       370       380

       140       150       160       170       180       190       200
  RATNMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGA
  | | || ||||| || ||| |||| ||  |||||||||| || || || ||    ||||| |||||||||
  GTTTAGAGAGTTGACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAGAA
390       400       410       420       430       440       450       460

     210       220       230       240       250       260       270
  RYTNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWS
  ||  |||   |  || ||||| ||  | || || ||||| ||||| |||  || ||||| || || || ||
  GCACTGCAAGGCCTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTC
       470       480       490       500       510       520       530

   280       290       300       310       320       330       340
  NAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCC
   |||||||||   || || |||   || ||||| ||||| |||||||| ||   |||| | ||||||| |
  TAAGATGCAAACGCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCT
     540       550       560       570       580       590       600

  350       360       370       380       390       400       410
  NCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGC
   ||    ||||| || ||||| ||||| ||||| || |  || ||  |||| || ||    ||||||  | |
  ACCAGAATGTCTGCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCTGCTTGG
    610       620       630       640       650       660       670

420       430       440       450       460       470       480
  NYTNAARCCNT------GGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWS
   || ||||| |      |||   |  | ||||||||||| || ||||| ||| | ||||||||||| |||||
  TCTGAAGCCCTGTATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTC
  680       690       700       710       720       730       740

    490       500       510       520       530                 540
  NWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNAC--NGCNCCNA------CNGCNC-CNCA
   || || || |  || ||  |||| || |   | || ||||| ||   || || |       | || |   ||
  CTCCACCCTGCTGCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCA
750       760       770       780       790       800       810       820

 550       560       570       580       590       600       610       620
  RCCNCCNYTNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNYTNYTNGCNGCNGCNTGGTGYYTNCAYTG
  ||     || || || || || || || ||  |     ||  | || || || || || ||| |||| | |||
  GCTGTTGCTCCTGCTGCTGCTGCTGCTGCCTCTCACACTGGTGCTGCTGGCAGCCGCCTGGGGCCTTCGCTG
       830       840       850       860       870       880       890

       630       640       650       660       670       680       690
  GCARMGNACNMGNMGNMGNACNCCNMGNCCNGGNGARCARGTNCCNCCNGTNCCNWSNCCNCARGAYYTNYT
  ||||||  | || || ||        |  ||       |   |  || || |  | || |  ||  |||||
  GCAAAGGGCAAGAAGGAGGGGGGAGC-TCCACCCTGGGGTGCCCCTCCCCTCCCAT--CCCTAGGATGCGAG
     900       910       920       930       940       950       960

     700    X
  NYTNGTNGARCAY
   || ||  | | |
  CCTTGTGCATCGTTGACTCAGCC
```

```
        970    X  980

5. US-08-162-407-6 (1-705)
   SCMIG1       S.cerevisiae MIG1 gene for a C2H2 zinc finger prot

 LOCUS       SCMIG1       2622 bp    DNA             PLN       23-JUL-1991
 DEFINITION  S.cerevisiae MIG1 gene for a C2H2 zinc finger protein involved in
             glucose repression.
 ACCESSION   X55734
 KEYWORDS    glucose repressor; MIG1 gene; zinc-finger protein.
 SOURCE      baker's yeast.
   ORGANISM  Saccharomyces cerevisiae
             Eukaryota; Plantae; Thallobionta; Eumycota; Hemiascomycetes;
             Endomycetales; Saccharomycetaceae.
 REFERENCE   1  (bases 103 to 2622)
   AUTHORS   Nehlin,J.O. and Ronne,H.
   TITLE     Yeast MIG1 repressor is related to the mammalian early growth
             response and Wilms' tumour finger proteins
   JOURNAL   EMBO J. 9, 2891-2898 (1990)
   STANDARD  full automatic
 REFERENCE   2  (bases 1 to 2622)
   AUTHORS   Ronne,H.E.
   TITLE     Direct Submission
   JOURNAL   Submitted (18-JUL-1991) to the EMBL/GenBank/DDBJ databases. H.E.
             Ronne, Ludwig Inst f Cancer Research, Uppsala Branch, Box 595,
             S-751 24 Uppsala, SWEDEN
   STANDARD  full automatic
 COMMENT     NCBI gi: 3962
 FEATURES             Location/Qualifiers
      source          1..2622
                      /organism="Saccharomyces cerevisiae"
                      /strain="W303-1A"
                      /haplotype="haploid"
                      /cell_type="MATa"
                      /clone_lib="pHR81"
                      /chromosome="VII"
                      /map="5.4 cM telomeric to TRP5"
      TATA_signal     780..785
      CDS             943..2457
                      /gene="MIG1"
                      /note="NCBI gi: 3963"
                      /codon_start=1
                      /product="glucose repressor"
                      /translation="MQSPYPMTQVSNVDDGSLLKESKSKSKVAAKSEAPRPHACPICH
                      RAFHRLEHQTRHMRIHTGEKPHACDFPGCVKRFSRSDELTRHRRIHTNSHPRGKRGRK
                      KKVVGSPINSASSSATSIPDLNTANFSPPLPQQHLSPLIPIAIAPKENSSRSSTRKGR
                      KTKFEIGESGGNDPYMVSSPKTMAKIPVSVKPPPSLALNNMNYQTSSASTALSSLSNS
                      HSGSRLKLNALSSLQMMTPIASSAPRTVFIDGPEQKQLQQQQNSLSPRYSNTVILPRP
                      RSLTDFQGLNNANPNNNGSLRAQTQSSVQLKRPSSVLSLNDLLVGQRNTNESDSDFTT
                      GGEDEEDGLKDPSNSSIDNLEQDYLQEQSRKKSKTSTPTTMLSRSTSGTNLHTLGYVM
                      NQNHLHFSSSSPDFQKELNNRLLNVQQQQQEQHTLLQSQNTSNQSQNQNQNQMMASSS
                      SLSTTPLLLSPRVNMINTAISTQQTPISQSDSQVQELETLPPIRSLPLPFPHMD"
 BASE COUNT      864 a    545 c    531 g    682 t
 ORIGIN

Initial Score    =    173  Optimized Score  =    368  Significance = 10.00
Residue Identity =     28%  Matches          =    341  Mismatches   =   322
Gaps             =     80  Conservative Substitutions               =     0
```



```
          X       10        20        30        40        50        60
          ATGACNGTNY-TNGCNCCNGCNTGGWSNCCNACNAC-NTAYYTNYTNYTNYTNYTNYTN
          | |   || |   |   | ||   |||      |  |   |||   | ||  | || |  ||
AGGAGAGTAAAAGCAAGTCCAAAGTAGCTGCGAAGTCAGAGGCGCCAAGACCACATGCTTGTCCTAT--CTG
      1010      1020      1030      1040      1050      1060        1070

        70        80        90       100       110       120       130
WSNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAAR
|| |   | || ||    ||   |||||| |    ||||   || |  || || |   ||   |  |  |||
TCATAGAGCTTTTC-----AC---AGACTGGA-ACATCAGACGAGACACATGAGAATTCATACAGGTGAGAAG
      1080              1090      1100      1110      1120      1130

     140       150       160       170       180       190
ATNMGNGARY-TNWSNGAY---TAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAY--
   |  | || | ||   |    || | |   ||| ||   | ||     |  || |  ||||  | ||||
CCTCACGCGTGTGACTTCCCCGGATGTGTGAAAAGGTT---CAGT---AGAAGCGATGAACTGACGAGACAC
  1140      1150      1160      1170              1180      1190      1200

200       210       220       230       240       250       260
-GARGARYT-NTGY---GGNGGNYTNTGGMGNY--TNGTNYTNGCNCARMGNTGGATG-GARMGNYTNA--A
 |||||||| | |         |  | ||| |     |     |    |||  || || | ||  |  |  |
AGAAGAATTCATACAAACTCCCACCCTCGAGGTAAAAGAGGCAGAAAGAAGAAGGTTGTGGGCTCTCCAATA
      1210      1220      1230      1240      1250      1260      1270

      270       280       290       300       310       320       330
RACNGTNGCNGGNWSNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTG
||  || ||  | ||   | ||| |     | |   |||   | ||||| | ||   |||||     |   ||
AATAGT-GCTAGTTC--TAGTGCTACCAGTATACCAGA-TTTAAATACGGCAA---ATTTTTCACCGCCATT
     1280      1290      1300      1310      1320          1330

     340       350       360       370       380       390
YGCNTTYCARCCNC---CNCC-NWSNTGYYT-NMGNT-TYG-TNCARACNAAYAT--NWSNMG-NYTNYTNC
  |    |||| |   | ||  |  | |||    | | ||| | | ||  || ||    || ||  || || |
ACCACAGCAACACCTATCGCCTTTAATTCCTATTGCTATTGCTCCGAAAGAAAATTCAAGTCGATCTTCTAC
1340      1350      1360      1370      1380      1390      1400

   400       410       420       430       440       450       460
ARGARACNWSNGARCARYTNGTNGCNYTNAARCCNTGGATNACNMGNCARAAYTTYWS-NMGNTGYYTNGAR
||||||     ||| |||          |  ||| |   ||  |   |  |||| | |  |  || ||
AAGAAAAGGTAGAAAAAC---CAAATTCGAAATCGGCGA-AAGTGGTGGGAATGACCCATATATGGTT---T
1410      1420          1430      1440      1450      1460      1470

  470       480       490       500       510       520       530
YTNCARTGYCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCN
|| |  | ||||   ||| | |  |  || ||  | |  || || || |  || |    ||| || |
CTTC--TCCCAAAACGATGGCT-AAGATTCCCGTCTCGGTGAAG-CCTCCACCTTCTTTAGC-ACTGAATAA
    1480      1490      1500      1510      1520      1530      1540

540       550       560       570       580       590       600
ACNG--CNCCNCARCCNCCNYTNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNY--TNYTNGCNGCNGC
   |  |  | |||  | |  || | || || |  || |  || |  | |  | |  | | || |  |
TATGAACTACCAAACTTCATCCGCTTCCACTGCTTTGTCTTC--GTTGAGCAATAGCCATAGTGGCAGTAGA
     1550      1560      1570      1580      1590      1600      1610

  610       620       630       640       650       660       670
NTGGTGYYTNCAYTGGCARMGNACNMGNMGNMGNACNCCNMGNCCNGGNGARCARGTNCCNCCNGTNCCNWS
 | |   ||  ||  | |    | |  |  |  |  | |  |   |  ||| ||| | |  |  |  |  ||
CT-GAAACTGAACGCGTTATCGTCCCTACAAATGATGACGC-CCATTGCTAGC-AGTGCGCCAAGGAC--TG
      1620      1630      1640      1650      1660      1670
```

```
    680         690         700     X
    NCCNCA-RGAY-YTNYT--NYTNGTNGARC-AY
        || ||||  | ||   |       ||| ||
    TTTTCATAGACGGTCCTGAACAGAAACAACTACAACAACAACA
  1680      1690      1700      1710      1720
```

6. US-08-162-407-6 (1-705)
   TAGLUIN1 Wheat Glu-Aly sublocus with Wis-2 insertion 5' end

```
 LOCUS       TAGLUIN1      794 bp    DNA             PLN       06-JUL-1989
 DEFINITION  Wheat Glu-Aly sublocus with Wis-2 insertion 5' end (0.8 kb Hind III
             fragment).
 ACCESSION   X05995
 KEYWORDS    Glu-Aly gene; glutenin; insertion sequence; retrotransposon;
             Wis-2 insertion sequence.
 SOURCE      wheat.
   ORGANISM  Triticum aestivum
             Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
             Commelinidae; Cyperales; Poaceae.
 REFERENCE   1  (bases 1 to 794)
   AUTHORS   Harberd,N.P., Flavell,R.B. and Thompson,R.D.
   TITLE     Identification of a transposon-like insertion in a Glu-1 allele of
             wheat
   JOURNAL   Mol. Gen. Genet. 209, 326-332 (1987)
   STANDARD  full automatic
 COMMENT     Sequence comparision with Cheyenne
             Glu-Aly sublocus gene has shown that the coding sequence of the
             Chinese Spring sublocus gene is interrupted by a 8.0 kb insertion
             sequence. A DNA duplication and inverted terminal sequences are
             flanking the insertion sequence. See also x05996.
             Data kindly reviewed (18-Apr-1988) by Flavell R.B.

             NCBI gi: 21780
 FEATURES             Location/Qualifiers
      source          1..794
                      /organism="Triticum aestivum"
                      /strain="Chinese spring"
                      /chromosome="chromosome 1a"
      misc_feature    <1..482
                      /note="glutenin homologous seq. (161 AA; includ. stop
                      codon) (1 is 2nd base in codon)"
      misc_feature    478..481
                      /note="DNA duplication"
      intron          483..>794
                      /note="insertion sequence"
      misc_feature    484..493
                      /note="inverted terminal sequence"
 BASE COUNT      290 a    183 c    178 g    139 t      4 others
 ORIGIN
```

```
Initial Score    =    164  Optimized Score  =    296  Significance =  9.36
Residue Identity =    25%  Matches          =    283  Mismatches   =   287
Gaps             =     77  Conservative Substitutions               =     0
```

```
     90       100       110       120       130           140       150
   YTGYWSNTTYCARC-AYWSNCCNATNWSNWSNGAYTTYGCNGTNA--AR-ATNMGNG-ARYTNWSNGAYTAY
              |||  |||  | |   | ||  |  | ||    |  || |  || | | |||  ||  |||||
            AAGCTTCTCCACAACAGCCAGGAAAATGGCAAGAACTAGGACAAGGGCAACAATGGTACTAT
            X       10        20        30        40        50        60
```



```
     160       170       180       190       200       210       220
 YTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNA-AYYTNCARGAYGARGARYTNTGYGGNGGNYTNTG--G
 |     ||  ||| | | |   | |  || || | |    ||| | | | || |||    ||      || ||  |
 C-----CA--ACTTCTCTG---CAGCAGCCAGGACAAGGGCAACAAG-GGTACT-ACCGAACTTCTCTGCAG
               70        80        90       100       110       120

      230       240       250       260       270       280       290
 MGNYTNGTNYTNGCNCARMGNTGG----ATGGARMGNYTNAARACNGTNGCNGGNWSNAARATGCARGGNYT
 |  |  |  |     ||||   ||     |  |||| ||   | | |  ||  | || |||  ||||
 CAGCCAGGACAAAGGCAACAAGGGTACTACCGAAC--TTCTCGGCAGCAGCCAGGAC-AAG-GGCAA--CAG
      130       140       150       160       170       180

       300       310       320       330       340       350       360
 NYTNGARMGNGTNAAY-ACNGARATNCAYTTYGTNACNAARTGYGCNTTYCA-RCCNCCNCCNWSNTGYYTN
  |  || |  |  ||| |  | | | | |  | |  | | | | || |   ||| | |     | ||   ||
 ATAGGACAATGGCAACAAGGGTACTACCCAACTTCTCCACA-GCAC---CCAGGACAAGGGCAAC-AACCAG
190       200       210       220       230       240       250

       370       380       390       400       410       420
 MGNTTYGTNCARACNAAYATNWSNM-GNYTNYTNCARGARACNWSNGARCARYTNGTNGCNYTNAARC--CN
  |    || |||| |||| || | |  |  | |||||   |  |||||| | | |  |||  | ||   |
 -GACAAGTGCAAA-AAATAGGACAAGGGCAACAACCAGAAAAAGGGCAACAAC---TAGGACAAGAGCAACA
    260       270       280       290       300       310       320

430       440       450       460       470       480       490
 TGGATNACNM-GNCAR-AAYTTYWS-NMGNTGYYTNGARYTNCARTGYCARCCNGAY-WSNWSNACNYTNCC
    | || | | ||| ||| |||  |  | |  ||| ||| | ||   || || || ||     |
 AATAGGACAAGGGCAACAAC--CAGAACAAGGGCAACAAC--CAGGACAAGGGCAACAACCAGGACAAGGGC
      330       340       350       360       370       380       390

 500       510       520       530       540       550       560
 NCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNAC-NGCNCCNCARCCNCCNYTNYTNYTNYTNY
   |    | ||  | |  |  ||    ||  | ||    ||  |  |  |||||   |  |  |  |  |
 AAC-AAGGGTACTACCCAAC--TTCTCTGCAGCAGCCAGGACAAGGGCAACAACCAGGACAATGACAACAAC
         400       410       420       430       440       450

570       580       590       600       610       620       630
 TNYTNY--TNCCNGTNGGNYTNYTNYT-NYTNGCNGCNGCNTGGTGYYTNCAYTGGCARMGNACNMGNMGNM
       |      |  | ||  ||          ||  | | | | |   ||    ||  |  |  |  |  |
 CAGGACAAGGGCAACAAGGGTACTGTTGGAAATATGC--CCTAGAGGC--AATAATAAATTGATTATTATTA
460       470       480       490       500       510       520

 640       650       660             670       680       690
 GNACNCCNMGNCCNGGNGARCARGT-----NCC--NCCNG---TNCCNWSNCCNCARGAYYTNYT--NYTNG
       ||  |  | |  ||  | |       ||   | |  |   ||    ||  ||  |  | |
 TATTTCCTNGTTCATGATAATCGTTTATTATCCATGCTAGAATTGTATTGATAGGAAACTCAGATACATGTG
 530       540       550       560       570       580       590

 700     X
 TNGA-RCAY
 | || ||||
 TGGATACATAGACAACACC
600       610
```

7. US-08-162-407-6 (1-705)
   TAGLU1AG Wheat gene for HMW-glutenin subunit located on chr

```
 LOCUS       TAGLU1AG     2915 bp    DNA             PLN       06-JUL-1989
```

```
DEFINITION  Wheat gene for HMW-glutenin subunit located on chromosome 1A.
ACCESSION   X03042
KEYWORDS    direct repeat; glutenin; inverted repeat; seed storage protein;
            signal peptide; storage protein.
SOURCE      wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
            Commelinidae; Cyperales; Poaceae.
REFERENCE   1  (bases 1 to 2915)
  AUTHORS   Forde,J., Malpica,J.M., Halford,N.G., Shewry,P.R., Anderson,O.D.,
            Greene,F.C. and Miflin,B.J.
  TITLE     The nucleotide sequence of a HMW glutenin subunit gene located on
            chromosome 1A of wheat (Triticum aestivum L.)
  JOURNAL   Nucleic Acids Res. 13, 6817-6832 (1985)
  STANDARD  full automatic
COMMENT     Though the premature stop codon may not affect gene expression the
            author suggests that the gene is a pseudogene, because no small in
            vivo transcript of chromosome 1A was found. Within the coding
            sequence are two multiple repeated amino acid motifs, PGQGQQ and
            GYYPTSLQQ
            Data kindly reviewed (19-MAR-1986) by J. Forde.

            NCBI gi: 21770
FEATURES             Location/Qualifiers
     source          1..2915
                     /organism="Triticum aestivum"
     repeat_unit     8..24
                     /note="inverted repeat X"
     repeat_region   243..259
                     /note="direct repeat B"
     repeat_region   272..291
                     /note="imp. direct repeat A"
     repeat_region   297..316
                     /note="imp. direct repeat A"
     repeat_region   318..337
                     /note="imp. direct repeat A"
     repeat_region   343..359
                     /note="direct repeat B"
     repeat_unit     419..431
                     /note="inverted repeat Y"
     promoter        582..586
                     /note="pot. CAAT-box"
     promoter        620..624
                     /note="pot. CAAT-box"
     repeat_unit     648..664
                     /note="inverted repeat X"
     promoter        696..699
                     /note="pot. CAAT-box"
     promoter        724..730
                     /note="put. TATA-box"
     CDS             815..2620
                     /note="precursor"
                     /codon_start=1
     CDS             815..877
                     /note="put. signal peptide (aa -21 to -1);  NCBI gi:
                     21771"
                     /codon_start=1
                     /translation="MAKRLVLFATVVIGLVSLTVA"
     CDS             878..2620
                     /note="mature HMW subunit of glutenin (aa 1-581)"
```



```
                     /codon_start=1
     misc_feature    1832..1834
                     /note="premature TAG stop codon"
     repeat_unit     2635..2647
                     /note="inverted repeat Y"
     misc_feature    2678..2683
                     /note="put. polyadenylation signal"
BASE COUNT      977 a    736 c    674 g    528 t
ORIGIN

Initial Score    =    164  Optimized Score  =    362  Significance =  9.36
Residue Identity =     27% Matches          =    344  Mismatches   =   326
Gaps             =     98  Conservative Substitutions              =     0

            X       10        20        30        40        50
            ATGACNGTNYTNGCNCCNGCNTGGWSNCCNACNA--CNTAYYT---NYTNYTNYTNYTNYTN
              |   |  |  ||  | ||    || || || |    |||||       |  || |
   CCTTCTACCATCGCGAGACCACGCCACTGCAGCAACTCCAACAAGGAATATTTGGGGGAACATCTTCACAAA
1110      1120      1130      1140      1150      1160      1170      1180

    60        70                  80        90       100       110
   YTNW-SNWSNGGNYT----NWSNGGNA------CNCARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNG
   |  | | || |  ||      || |  |      | ||| |  | || | | | | ||| || |  |  ||
   CAGTACAAGGGTATTACCCAAGTGTAATATCTCCTCAGCAGGGGTCATATTATC-CAGGCCAAGCTTCTC-C
        1190      1200      1210      1220      1230      1240      1250

  120       130       140       150       160       170       180
   AYTTY-GCNGTNAARAT-NMGNGARYTNWSNGAYTA--YYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSN
   ||  | ||    |||||  |  ||||| || ||| |   |  |    | |||||||     |  | || ||
   ACAACAGCCAGGAAAATGGCAAGAACT-AG-GACAAGGGCAACAATGGTACTATCCAACTTCTCT-GC-AGC
        1260      1270      1280      1290      1300      1310

    190           200       210       220           230       240
   AAYY--TNCARG---AYGARGARYTNT-GYGGNGGNYTNTG--GMGNYTNGTNYTNGCNCARMGNTGG----
   | ||    ||||   |  ||| ||  |  ||     || ||  ||  |  |  |     ||||   ||
   AGCCAGGACAAGGGCAACAAGGGTACTACCGAACTTCTCTGCAGCAGCCAGGACAAAGGCAACAAGGGTACT
1320      1330      1340      1350      1360      1370      1380      1390

    250       260       270           280       290           300       310
   ATGGARMGNYTNAARACNGTNGCNGGNWSNA----ARATGCARGGNYTNYTNGA--RMGNGTNAAYACNGAR
   |  |||| ||      | |  ||  | || |    ||  |  ||| |  |   |  ||| || | |||  ||
   ACCGAAC--TTCTCTGCAGCAGCCAGGACAAGGGCAACAGATAGGACAATGGCAACAAGGGT-ACTACCCAA
         1400      1410      1420      1430      1440      1450      1460

       320       330       340       350       360       370       380
   ATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSN
   | | |  |   ||  || |       ||||  ||    ||  | |  ||   ||  ||||  ||||  ||
   CTTCTCCGCAGCACCCAG-GACAAGGGCAACAACC----AG--GAC-AAG---TG--CAAA-AAATAGGACA
         1470      1480      1490           1500           1510

      390       400       410           420           430       440
   M-GNYTNYTNCARGARACNWSNGARCARYTNG-----TNGCNYTNAA---RCCNTGGATNACNMGNCARAAY
   | | |  | ||||   |  |||||| |        ||    ||   | |  ||   || |  ||||||
   AGGGCAACAACCAGAAAAAGGGCAACAACTAGGACAAGAGCAACAAATAGGACAAGGGCAAC-AACCAGAAC
1520      1530      1540      1550      1560      1570      1580

   450       460       470       480       490       500       510
   TTYWSNMGNTGYYTNGARYTNCARTGY--CARCC-NGAYW----SNWSNACNYTNCCNCCNCCNTGGWSNCC
      | |    ||  ||    || || ||||  ||||    | || |   | |  ||    |  ||  |
   AAGGGCAACAACC---AGGACAAGGGCAACAACCAGGACAAGGGCAACAAGGGTACTACCCAACTTCTCTGC
```

```
1590      1600         1610      1620      1630      1640      1650

        520           530       540       550       560       570
   NMGNCCNYTNGA----RGCNACNGCNCCNAC-NGCNCCNCARCCNCCNYTNYTNYTNYTNYTNYTNCCN
    ||    |  ||    ||| ||  |    ||      |  |||||    |      |  |       | || ||
   -AGCAGCCAGGACAAGGGCAACAACCAGGACAATGGCAACAACCAGGACAAGGGCAAC-AAGGGTACTACC-
  1660      1670      1680      1690      1700      1710      1720

 580       590       600       610       620       630       640
   GTNGGNYTNYTNYTNYTNGCNGCNGCNTGGTGYYTNCAYTGGCA--RMGNACNMGNMGNMGNA-CNCCNMGN
         || || || |       | |       | |  ||  ||||  ||  |  |  |  || | |  | ||
   ---CAACTTCT-CTGCAACAGCCAGTACAAGGGCAACAAGGGCACTACCCA-GCTTCTCAGCACCAGCCAGG
     1730      1740      1750      1760      1770      1780      1790

  650       660       670       680       690       700    X
   CCNGGNG-ARCARGTNCCNCCNGTNCCNWSNCCNCARGAYYTNYTNYTNGTNGARCAY
    | || | ||||||  |  |       | || |  ||| | |      |  |   ||||
   GCAGGGGCAACAAGGGCACCAGCCAGCTTCTCTGCAGTAGTCAGGACAAGGGCAACAAGGGCACCACC
     1800      1810      1820      1830      1840      1850 X    1860
```

8. US-08-162-407-6 (1-705)
HSCOL16A Homo sapiens alpha-1 type XVI collagen (COL16A1) m

```
ID   HSCOL16A   standard; RNA; PRI; 5387 BP.
XX
AC   M92642;
XX
DT   30-APR-1992 (Rel. 31, Created)
DT   06-JAN-1995 (Rel. 42, Last updated, Version 2)
XX
DE   Homo sapiens alpha-1 type XVI collagen (COL16A1) mRNA, complete
DE   cds.
XX
KW   alpha-1 type XVI collagen; extracellular matrix protein.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC   Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
XX
RN   [1]
RP   1-5387
RA   Pan T.C., Zhang R.Z., Mattei M.G., Timpl R., Chu M.L.;
RT   "Cloning and chromosomal location of human alpha 1(XVI) collagen";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).
XX
CC   NCBI gi: 180757
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..5387
FT                   /organism="Homo sapiens"
FT                   /cell_type="Fibroblasts"
FT                   /sequenced_mol="cDNA to mRNA"
FT   5'UTR           1..153
FT                   /gene="COL16A1"
FT   sig_peptide     154..216
FT                   /gene="COL16A1"
FT                   /codon_start=1
FT   CDS             154..4965
```



```
FT                   /gene="COL16A1"
FT                   /note="NCBI gi: 180758"
FT                   /codon_start=1
FT                   /product="alpha-1 type XVI collagen"
FT   mat_peptide     217..4962
FT                   /partial
FT                   /gene="COL16A1"
FT                   /evidence=EXPERIMENTAL
FT                   /codon_start=1
FT                   /product="alpha-1 type XVI collagen"
FT   3'UTR           4987..5387
FT                   /gene="COL16A1"
XX
SQ   Sequence 5387 BP; 1153 A; 1520 C; 1817 G; 897 T; 0 other;

Initial Score    =    160  Optimized Score  =    306  Significance =  9.08
Residue Identity =    27%  Matches          =    276  Mismatches   =   296
Gaps             =     62  Conservative Substitutions               =     0

   100       110       120       130       140       150       160       170
    RCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARATNMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCC
                  ||  | |||  |   |   |||| || ||||     |   ||||     |||| |||
               GGGACCCGTGACCGTCACGGAGAGAT-AGAGAACCCCTAGCGGACCT----GAGGA--ACCT
               X       10        20        30        40            50

          180       190       200       210       220       230       240
    NGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYTNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMG
     |   | |   | ||   ||| | |  || |  |||  || | ||      ||| |  | |  |   |  ||
    GGGCTCTGCTTCCTC-CGCCTGCTGCCTGGACTGCTCAGT-GAGGTGGGGGGATCCCCG-GCCACCGGTGCT
       60        70        80        90        100       110       120

        250       260       270       280       290       300
    NTGGATGGARMGNYTNAARACNGTNG--CNGGNWSNAAR-ATGCARGGNYTNYTNGARMGNGTNAAYAC---
     ||| || ||   || |   |  | | | || ||   | || | ||| || || |   | |    | |
    GTGGCTGTAG---CTGACCTCTTTGGACCGGGATGTGGGTATCCTGGG-CTCCTGGCCTGTGGCTGCTCGGT
       130           140       150       160       170       180       190

      310       320           330       340       350       360
    ---NGARATNCAYTTYGTNA----CNAARTGYGCNTTY-CARCCNCCNCCNWS-NTGYYTNM--GNTTY--G
        | | | ||  |  |       |||| |   | | |||   || || ||   ||   |  |  ||
    CTTTGGGCTACCTTCGGCCATGGGGCAAATACAGGTGCACAATGCCCACCTTCACAGCAGGAAGGACTCAAA
         200       210       220       230       240       250       260

    370       380       390       400       410       420       430
    TNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGCNYTNAARCCNTGGATNA--C
    |   |||  |||  || || || |  |||  |||  | ||   |||  |    || |  |   |  | | |
    TTGGAAC--ACAGTAGTAGCCTGC--CAGCCAAC--GTGACTGGCT--TCAACCTCATCCACCGACTCAGCC
       270       280       290       300       310       320

    440       450       460       470       480           490       500
    NMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAY------WSNWSNACNYTNCCNCCNCCN
     |   || |  |||| |       || |  |||  ||||   | |      || ||  | |     |  ||
    TCATGAAGAAGTCTGCCATCAAGAAGATCCGCAACCCCAAGGGGCCTCTCATCCTGCGCCTGGGGGCGGCCC
   330       340       350       360       370       380       390       400

       510       520       530       540       550       560       570
    TGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCNCARCCNCCNYT--NYTNYTNYTNYTNYTNY
      ||| || || || |  || |     | |     |   | | |      ||   ||  | ||    || |
    CCGTGACCCAGCCCACGCGACGAGTATTCCCTCGGGGTCTCCCGGAGGAGTTTGCCCTGGTGCTGACACTAC
          410       420       430       440       450       460       470
```

```
      580       590       600           610       620       630       640
TNCCNGTNGGNYTNYTNYTNYTNGCNGCNGCNTGG----TGYYTNCAYTGGCARM-GNACNMGNMGNMGNAC
|  | |  |     |  |  |  |  |   | |||    |||||  | || | || | |  |       ||
T-GCTGAAGAAACACACCCACCAGAAG--ACGTGGTATCTGTTTCAAGTGACCGATGCAAATGGGTATCCAC
      480       490       500       510       520       530       540

      650           660       670       680       690       700     X
NCCNMGNCCNG---GNGARCARGTNCCNCCNGTNCCNWSNCCNCARGAYYTNYTNYTNGTNGARC-AY
    |  || |   |  |||||       | |  ||  |  | |||| ||     |  |    || ||
AGATATCCCTGGAAGTCAACAGCCAAGAGCGGAGCCTGGAGCTCAGGGCCCAGGGCCAGGATGGCGACTTTG
      550       560       570       580       590       600       610

TGTCCT
```

9. US-08-162-407-6 (1-705)
SCBYP13CS S.cerevisiae gene for byp1-3 complementing sequenc

```
LOCUS       SCBYP13CS    4100 bp    DNA             PLN       24-NOV-1993
DEFINITION  S.cerevisiae gene for byp1-3 complementing sequence.
ACCESSION   X55442
KEYWORDS    byp1-3 complementing sequence.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Plantae; Thallobionta; Eumycota; Hemiascomycetes;
            Endomycetales; Saccharomycetaceae.
REFERENCE   1  (bases 1 to 4100)
  AUTHORS   Huse,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-1990) to the EMBL/GenBank/DDBJ databases. Huse
            K., Institut fuer Mikrobiologie, TH Darmstadt, Schnittspahnstr. 10,
            D-6100 Darmstadt
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 4100)
  AUTHORS   Huse,K., Hohmannn,S., Valentin,E. and Zimmermann,F.K.
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   3  (bases 3841 to 4100)
  AUTHORS   Robison,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1993) to the EMBL/GenBank/DDBJ databases. Sent to
            EMBL by K. Robison, email:robison@biosun.harvard.edu
  STANDARD  full automatic
COMMENT     NCBI gi: 3436
FEATURES             Location/Qualifiers
     source          1..4100
                     /organism="Saccharomyces cerevisiae"
                     /strain="AB 320"
                     /chromosome="VII"
     CDS             2680..3840
                     /note="NCBI gi: 3437"
                     /codon_start=1
                     /product="MIG1"
                     /translation="MQSPYPMTQVSNVDDGSLLKESKSKSKVAAKSEAPRPHACPICH
                     RAFHRLEHQTRHMRIHTGEKPHACDFPGCVKRFSRSDELTRHRRIHTNSHPRGKRGRK
                     KKVVGSPINSASSSATSIPDLNTANFSPPLPQQHLSPLIPIAIAPKENSSRSSTRKGR
                     KTKFEIGESGGNDPYMVSSPKTMAKIPVSVKPPPSLALNNMNYQTSSASTALSSLSNS
                     HSGSRLKLNALSSLQMMTPIASSAPRTVFIDGPEQKQLQQQQNSLSPRYSNTVILPRP
```



```
                     RSLTDFQGLNNANPNNNGSLRAQTQSSVQLKRPSSVLSLNDLLVGQRNTNESDSDFTT
                     GGEDEEDGLKDPSNSSIDNLEQDYLQEQSRKKSKTSTPTTMLSRSTGRVWSP"
     misc_feature    3841..4100
                     /note="putative VECTOR sequence pBR322"
                     /citation=[3]
BASE COUNT     1249 a    821 c    882 g   1148 t
ORIGIN

Initial Score    =    156  Optimized Score  =    368  Significance =  8.79
Residue Identity =    28%  Matches          =    341  Mismatches   =   322
Gaps             =     80  Conservative Substitutions               =     0

             X       10        20        30        40        50        60
             ATGACNGTNY-TNGCNCCNGCNTGGWSNCCNACNAC-NTAYYTNYTNYTNYTNYTNYTN
             | |   || |   |   | ||   |||     |  |   |||   | ||  | || |  ||
   AGGAGAGTAAAAGCAAGTCCAAAGTAGCTGCGAAGTCAGAGGCGCCAAGACCACATGCTTGTCCTAT--CTG
  2740      2750      2760      2770      2780      2790      2800

          70        80        90       100       110       120       130
   WSNWSNGGNYTNWSNGGNACNCARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAAR
   || |   | || ||    ||  |||||| |    ||||   || |  || || |   ||   |  | |||
   TCATAGAGCTTTTC-----AC---AGACTGGA-ACATCAGACGAGACACATGAGAATTCATACAGGTGAGAAG
  2810      2820           2830      2840      2850      2860      2870

       140       150       160       170       180       190
   ATNMGNGARY-TNWSNGAY---TAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAY--
     |  | || | ||   |   || | |   ||| ||   | ||    |  || |  |||| | ||||
   CCTCACGCGTGTGACTTCCCCGGATGTGTGAAAAGGTT---CAGT---AGAAGCGATGAACTGACGAGACAC
       2880      2890      2900      2910      2920      2930

   200       210       220       230       240       250       260
   -GARGARYT-NTGY---GGNGGNYTNTGGMGNY--TNGTNYTNGCNCARMGNTGGATG-GARMGNYTNA--A
    |||||||| | |       | | ||| |    |     |    |||  || || | ||  |  |  |
   AGAAGAATTCATACAAACTCCCACCCTCGAGGTAAAAGAGGCAGAAAGAAGAAGGTTGTGGGCTCTCCAATA
  2940      2950      2960      2970      2980      2990      3000

          270       280       290       300       310       320       330
   RACNGTNGCNGGNWSNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTG
   ||  || ||  | ||   | ||| |    | |   |||   | ||||| | ||   |||||    |   ||
   AATAGT-GCTAGTTC--TAGTGCTACCAGTATACCAGA-TTTAAATACGGCAA---ATTTTTCACCGCCATT
3010      3020      3030      3040      3050      3060      3070

       340       350       360       370       380       390
   YGCNTTYCARCCNC---CNCC-NWSNTGYYT-NMGNT-TYG-TNCARACNAAYAT--NWSNMG-NYTNYTNC
     |    ||||  |   | ||  |  | |||   | | ||| | | ||  || ||   || ||  || || |
   ACCACAGCAACACCTATCGCCTTTAATTCCTATTGCTATTGCTCCGAAAGAAAATTCAAGTCGATCTTCTAC
     3080      3090      3100      3110      3120      3130      3140

     400       410       420       430       440       450       460
   ARGARACNWSNGARCARYTNGTNGCNYTNAARCCNTGGATNACNMGNCARAAYTTYWS-NMGNTGYYTNGAR
   ||||||     ||| |||         |  ||| |   ||  |   |  ||||  | |  |  || ||
   AAGAAAAGGTAGAAAAAC---CAAATTCGAAATCGGCGA-AAGTGGTGGGAATGACCCATATATGGTT---T
   3150      3160      3170      3180      3190      3200      3210

     470       480       490       500       510       520       530
   YTNCARTGYCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCN
   || |  | ||||   ||| | |  |  || ||  | |  || || || |  || |    ||| || |
   CTTC--TCCCAAAACGATGGCT-AAGATTCCCGTCTCGGTGAAG-CCTCCACCTTCTTTAGC-ACTGAATAA
         3220      3230      3240      3250      3260      3270
```

```
   540         550         560         570         580         590         600
   ACNG--CNCCNCARCCNCCNYTNYTNYTNYTNYTNYTNCCNGTNGGNYTNYTNY--TNYTNGCNGCNGC
      |  |  |  |||  |  |  || |  || || |  || |  || |  |  |  |  |  | || |  |
   TATGAACTACCAAACTTCATCCGCTTCCACTGCTTTGTCTTC--GTTGAGCAATAGCCATAGTGGCAGTAGA
 3280       3290        3300        3310        3320        3330        3340

     610         620         630         640         650         660         670
   NTGGTGYYTNCAYTGGCARMGNACNMGNMGNMGNACNCCNMGNCCNGGNGARCARGTNCCNCCNGTNCCNWS
    | |   ||  ||  |  |    | |  |  |  |   | |  |   |  ||| ||| |  |  |  |  ||
   CT-GAAACTGAACGCGTTATCGTCCCTACAAATGATGACGC-CCATTGCTAGC-AGTGCGCCAAGGAC--TG
 3350        3360        3370        3380        3390        3400        3410

   680         690          700     X
   NCCNCA-RGAY-YTNYT--NYTNGTNGARC-AY
       || ||||  | ||   |       ||| ||
   TTTTCATAGACGGTCCTGAACAGAAACAACTACAACAACAACA
    3420        3430        3440        3450


10. US-08-162-407-6 (1-705)
    WHTGLIABF    Wheat (T.aestivum) alpha-/beta-gliadin class A-I m

 LOCUS       WHTGLIABF     950 bp ss-mRNA            PLN       02-MAY-1986
 DEFINITION  Wheat (T.aestivum) alpha-/beta-gliadin class A-I mRNA, complete
             cds, clone pA26.
 ACCESSION   M11074
 KEYWORDS    gliadin.
 SOURCE      Wheat (T.aestivum, cv. Cheyenne) seed, cDNA to mRNA, clone pA26.
   ORGANISM  Triticum aestivum
             Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
             Commelinidae; Cyperales; Poaceae.
 REFERENCE   1  (bases 1 to 950)
   AUTHORS   Okita,T.W., Cheesbrough,V. and Reeves,C.D.
   TITLE     Evolution and heterogeneity of the alpha-/beta-type and gamma-type
             gliadin DNA sequences
   JOURNAL   J. Biol. Chem. 260, 8203-8213 (1985)
   STANDARD  full automatic
 COMMENT     Draft entry and sequence in computer readable form for [1] kindly
             provided by T.W.Okita, 25-OCT-1985.
             The alpha-/beta-gliadins can be divided into five homology classes.
             Sequence divergence between the classes is due to single base
             substitutions and to duplications or deletions within or near
             direct repeats.  There are more than a hundred copies of the gene
             encoding alpha-/beta-gliadin per haploid genome.  Amplification or
             rearrangement of DNA does not occur during development, since
             hybridization patterns are the same, whether DNA isolated from
             leaves or seeds is used.

             NCBI gi: 170721
 FEATURES             Location/Qualifiers
      source          1..950
                      /organism="Triticum aestivum"
      mRNA            <1..>950
                      /note="gliadin mRNA"
      sig_peptide     49..108
                      /note="alpha-/beta-gliadin signal peptide"
                      /codon_start=1
      CDS             49..837
                      /note="pre-alpha-/beta-gliadin A-I;  NCBI gi: 170722"
                      /codon_start=1
```



```
                      /translation="MKTFLILALLAIVATTATTAVRVPVPQLQPQNPSQQQPQEQVPL
                      VQQQQFLGQQQPFPPQQPYPQPQPFPSQQPYLQLQPFLQPQLPYSQPQPFRPQQPYPQ
                      PQPQYSQPQQPISQQQQQQQQQQQQQQQQQQQIIQQILQQQLIPCMDVVLQQHNIVHG
                      KSQVLQQSTYQLLQELCCQHLWQIPEQSQCQAIHNVVHAIILHQQQKQQQQPSSQVSF
                      QQPLQQYPLGQGSFRPSQQNPQAQGSVQPQQLPQFEEIRNLARK"
      mat_peptide     109..834
                      /note="alpha-/beta-gliadin A-I"
                      /codon_start=1
 BASE COUNT      321 a    296 c    146 g    187 t
 ORIGIN      26 bp upstream of AluI site; chromosome 6.

Initial Score    =    153  Optimized Score  =    227  Significance =  8.58
Residue Identity =    28%  Matches          =    203  Mismatches   =   206
Gaps             =     41  Conservative Substitutions               =     0

  260       270         280         290         300         310         320         330
   ARACNGTNGCNGGNWSNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAART
              |   | |  ||  |||     |  |    || |||||  |||| |||  ||     |||
             AGAGCGCATCATTTAAGCCAAGCAAGCTGTGGTCAATAC--AAATCCACCATG-----AAA-
            X         10          20          30          40          50

          340         350         360         370         380         390         400
   GYGCNTTYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGAR-ACN
    |   | ||| ||    ||  | ||||     ||||   ||| | |   || |  || |    ||||  |
   ACCTTTCTCATCCTTGCCCTCCTTGCT-----ATCGTGGCGACCACCGCCACAA--CTGCAGTTAGAGTTCC
       60          70          80          90          100         110

          410         420         430         440         450         460         470
   WSNG-ARCARYTNGTNGCNYTNAARCCNT-GGATNACNMGNCARAAYTTYWSNMGNTGY-YTNGARYTNCAR
   || |  ||||||     | |  || || |   |  |  || || ||   ||  || | | || | ||  |||
   AGTGCCACAATTGCAGCCACAAAATCCATCTCAGCAACAGCCACAAGAGCA--AGTTCCATTGGTACAACAA
 120         130         140         150         160         170         180

          480         490         500         510         520         530
   TGYCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNAC----
      ||| |   ||| ||    |  |  |  || |   | || |  |  |    | | || ||  | ||
   ---CAA-CAATTTCTAG--GGCAGCAACAACCATTTCCACCACAACAAC-CATATCCACAGCCGCAACCATT
    190         200         210         220         230         240         250

 540       550         560         570         580         590         600
   NGCNCCNCARCCNCCNYTNYTNYTNYTNYTNYTNC--CNGTNGGNYTNYTNY--TNYTNGCNGCNGC-N
     |  | ||||  || |  || | || |  |  || |  | |  |  |  |  || |  | ||  | ||  |
   TCCATCACAACAACCATATCTGCAACTGCAACCATTTCTGCAGCCG--CAACTACCATATTCACAGCCACAA
      260         270         280         290         300         310         320

  610         620         630         640         650         660         670
   TGGTGYYTNC-AYTGGCARMGNACNMGNMGNMGNACNCCNMGNCCNGGNGARCARGTNCCNCCNGTNCCNWS
      | ||  | ||    | |  |  |  |  || |  |  ||     |  ||| |   |  ||  |  |
   CCATTTCGACCACAACAACCATATCCACAACCGCAACCACAGTATTCG-CAACCACAACAACCAATTTCACA
      330         340         350         360         370         380         390

   680       690       700    X
   NCCNCARGAYYTNYTNYTNGTNGARCAY
    |  ||| | |  |  |      ||||
   GCAGCAGCAGCAGCAGCAACAACAACAACAAC
     400       410       420  X    430
```

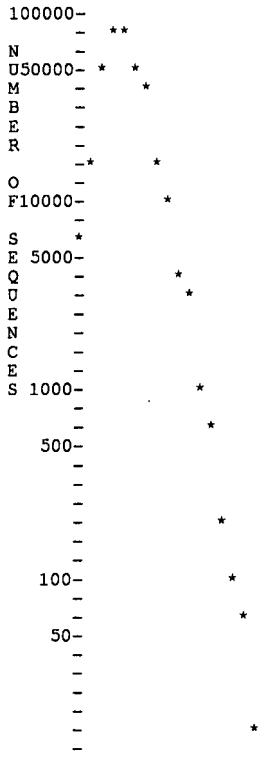
maryh@stic **stdin**

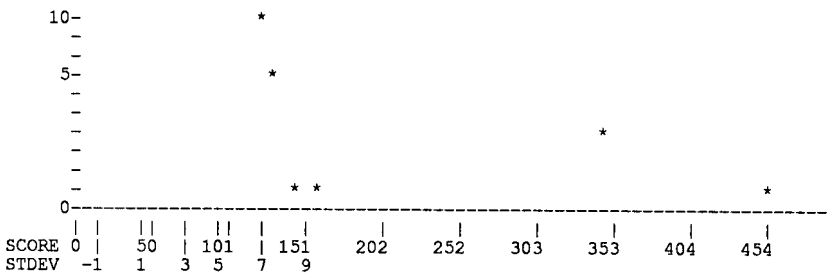
NeWSprinter20

Thu Apr 06 10:26:39 1995

NeWSprint 2.5 Rev B
Openwin library 3
NeWSprint interpreter 210.0

```
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file transl6.res made by maryh on Thu 6 Apr 95 9:48:27-PDT.


Query sequence being compared:US-08-162-407-6 (1-465)
Number of sequences searched:               313646
Number of scores above cutoff:                4147

      Results of the initial comparison of US-08-162-407-6 (1-465) with:
   Data bank : EMBL-NEW 1, all entries
   Data bank : GenBank 86, all entries
   Data bank : GenBank-NEW 1, all entries
   Data bank : UEMBL 41_86, all entries
```





PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	32	31	13.51

Times:	CPU	Total Elapsed
	00:27:06.00	00:27:22.00

Number of residues: 288682119
Number of sequences searched: 313646
Number of scores above cutoff: 4147

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 31 standard deviations above mean ****					
1. HSU03858	Human flt3 ligand mRNA, compl	1080	454	465	31.23	0
2. HSU04806	Human FLT3/FLK2 ligand mRNA,	859	452	464	31.08	0
	**** 23 standard deviations above mean ****					
3. MUSLIGAND	Mouse flt3 ligand mRNA, compl	829	345	400	23.16	0
4. MMU04807	Mus musculus FLT3/FLK2 ligand	1152	343	399	23.01	0
	**** 9 standard deviations above mean ****					
5. MZEPOD	Maize pyruvate,orthophosphate	3171	161	241	9.55	0
	**** 8 standard deviations above mean ****					
6. HHU06941	Human herpesvirus 6B R1 repli	1651	145	240	8.36	0
	**** 7 standard deviations above mean ****					

7. TAGLUIN1	Wheat Glu-Aly sublocus with W	794	134	228	7.55	0
8. COTMYBA	Cotton DNA-binding domain mRN	1006	134	249	7.55	0
9. SAUNKORFB	S.ambofaciens ORF, complete C	1449	134	242	7.55	0
10. TAGLU1AG	Wheat gene for HMW-glutenin s	2915	134	241	7.55	0
11. MMIGE6H	Mouse mRNA for anti-hPLAP-dir	1579	128	252	7.10	0
12. MDPMVT10	Aleutian mink disease parvovi	1601	127	239	7.03	0
13. MDPCGA	Aleutian mink disease parvovi	4801	127	239	7.03	0
	**** 6 standard deviations above mean ****					
14. BLYSEQ	Hordeum vulgare DNA sequence	830	126	244	6.96	0
15. DROOTUA	D.melanogaster ovarian tumor	5161	126	249	6.96	0
16. HSHZF6	H.sapiens HZF6 mRNA for zinc	627	124	227	6.81	0
17. SCYBR216C	S.cerevisiae chromosome II re	2517	123	238	6.73	0
18. MZEPPDKA2	Corn pyruvate,orthophosphate	5519	122	239	6.66	0
19. CHPSZFPA	Pea chloroplast accD gene for	1254	121	241	6.59	0
20. CHPSZFPG	Pisum sativum chloroplast orf	3240	121	241	6.59	0

1. US-08-162-407-6 (1-465)
HSU03858 Human flt3 ligand mRNA, complete cds.

```
LOCUS       HSU03858     1080 bp    mRNA            PRI       19-JUL-1994
DEFINITION  Human flt3 ligand mRNA, complete cds.
ACCESSION   U03858
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Lyman,S.D., James,L., Johnson,L., Brasel,K., de Vries,P.,
            Escobar,S.S., Downey,H., Splett,R.R., Beckmann,M.P. and
            McKenna,H.J.
  TITLE     Cloning of the human homologue of the murine flt3 ligand: a growth
            factor for early hematopoietic progenitor cells
  JOURNAL   Blood 83, 2795-2801 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Lyman,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-1993) Stewart D. Lyman, Immunex Research and
            Development Corporation, 51, University St., Seattle, WA 98101, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 494978
FEATURES             Location/Qualifiers
     source          1..1080
                     /clone_lib="expression, cDNA"
                     /organism="Homo sapiens"
                     /cell_line="clone 22 (T cell)"
     5'UTR           1..83
     CDS             84..791
                     /standard_name="FMS-like tyrosine kinase-3 ligand"
                     /note="ligand for the flt3/flk-2 tyrosine kinase receptor;
                      NCBI gi: 494979"
                     /codon_start=1
                     /function="stimulates proliferation of early hematopoietic
                     cells"
                     /product="flt3 ligand"
                     /translation="MTVLAPAWSPTTYLLLLLLLSSGLSGTQDCSFQHSPISSDFAVK
                     IRELSDYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTVAGSKMQGLLERVN
                     TEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQP
```



```
                     DSSTLPPPWSPRPLEATAPTAPQPPLLLLLLLPVGLLLLAAAWCLHWQRTRRRTPRPG
                     EQVPPVPSPQDLLLVEH"
     sig_peptide     84..161
     misc_feature    162..629
                     /note="extracellular domain"
     misc_feature    630..698
                     /note="transmembrane domain"
     misc_feature    699..788
                     /note="cytoplasmic domain"
     3'UTR           792..1080
     misc_feature    1015..1080
                     /note="ATTTA mRNA instability motif"
     polyA_signal    1059..1064
     polyA_site      1080
                     /note="32 A residues"
BASE COUNT      204 a    384 c    290 g    202 t
ORIGIN
```

```
Initial Score    =    454  Optimized Score  =    465  Significance = 31.23
Residue Identity =     55% Matches          =    370  Mismatches   =    95
Gaps             =      0  Conservative Substitutions               =     0

            X       10        20        30        40        50        60
            CARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARATNMGNGARYT
            |||||||||||| |||||||||| || || || || |||||||| || ||||| || |||||
  CAGTGGGACCCAGGACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCT
     160    X 170       180       190       200       210       220

         70        80        90       100       110       120       130
  NWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYTNTGYGGNGG
   || |||||||| || |||||||||| || || || || || ||||| |||||||||||||| ||||| ||
  GTCTGACTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGG
   230       240       250       260       270       280       290

     140       150       160       170       180       190       200
  NYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNAARATGCARGG
   || ||||| || || || || ||||| |||||||||| || ||||| || || || || ||||||||||
  CCTCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGG
300       310       320       330       340       350       360       370

   210       220       230       240       250       260       270
  NYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCNCCNWSNTGYYT
   || || ||||| || ||||| ||||| |||||||| || |||||||| |||||||| || || || |||||
  CTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCT
         380       390       400       410       420       430       440

280       290       300       310       320       330       340       350
  NMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGCNYTNAARCCNTG
   || ||||| ||||| ||||| || || || || |||||||| || |||||||| |||||||| || || |||||
  TCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTG
       450       460       470       480       490       500       510

          360       370       380       390       400       410       420
  GATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWSNWSNACNYTNCCNCCNCC
  ||| || || |||||||||| || ||||| ||||| |||||||||| ||||| || || || || || ||
  GATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCACCCCC
     520       530       540       550       560       570       580

       430       440       450       460    X
  NTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCN
```

```
     ||||| || || || || ||||| || || || || || ||
    ATGGAGTCCCCGGCCCCTGGAGGCCACAGCCCCGACAGCCCCGCAGCCCCCTC
     590       600       610       620       630
```

2. US-08-162-407-6 (1-465)
   HSU04806 Human FLT3/FLK2 ligand mRNA, complete cds.

```
 LOCUS       HSU04806      859 bp    mRNA            PRI       11-MAY-1994
 DEFINITION  Human FLT3/FLK2 ligand mRNA, complete cds.
 ACCESSION   U04806
 KEYWORDS    .
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 798)
   AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
             Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
             Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
             Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
             Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
   TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
             haematopoietic stem cells and is encoded by variant RNAs
   JOURNAL   Nature 368, 643-648 (1994)
   STANDARD  full automatic
 REFERENCE   2  (bases 1 to 859)
   AUTHORS   Culpepper,J.A.
   TITLE     Direct Submission
   JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
             Institute, 901 California Avenue, Palo Alto, CA 94304, USA
   STANDARD  full automatic
 COMMENT     NCBI gi: 483844
 FEATURES             Location/Qualifiers
      source          1..859
                      /clone="human Flt3/Flk2 ligand S86"
                      /clone_lib="SV48 cDNA"
                      /organism="Homo sapiens"
                      /cell_line="SV48 thymic stroma"
      CDS             93..800
                      /note="NCBI gi: 483845"
                      /codon_start=1
                      /product="FLT3/FLK2 ligand"
                      /translation="MTVLAPAWSPTTYLLLLLLLSSGLSGTQDCSFQHSPISSDFAVK
                      IRELSDYLLQDYPVTVASNLQDEELCGALWRLVLAQRWMERLKTVAGSKMQGLLERVN
                      TEIHFVTKCAFQPPPSCLRFVQTNISRLLQETSEQLVALKPWITRQNFSRCLELQCQP
                      DSSTLPPPWSPRPLEATAPTAPQPPLLLLLLLPVGLLLLAAAWCLHWQRTRRRTPRPG
                      EQVPPVPSPQDLLLVEH"
 BASE COUNT      147 a    318 c    246 g    148 t
 ORIGIN
```

```
Initial Score    =    452  Optimized Score  =    464  Significance = 31.08
Residue Identity =    55%  Matches          =    369  Mismatches   =    96
Gaps             =      0  Conservative Substitutions               =     0

             X       10        20        30        40        50        60
             CARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARATNMGNGARYT
             |||||||||||| |||||||||| || || || || |||||||| || ||||| || |||||
    CAGTGGGACCCAGGACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCT
       170   X   180       190       200       210       220       230
```



```
          70        80        90       100       110       120       130
    NWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYTNTGYGGNGG
     || |||||||| || |||||||||||| || || || || || ||||| |||||||||||||| ||||| |
    GTCTGACTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGC
      240       250       260       270       280       290       300

        140       150       160       170       180       190       200
    NYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNAARATGCARGG
     || ||||| || || || || ||||| |||||||||||| || ||||| || || || || ||||||||||
    GCTCTGGCGGCTGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAGG
    310       320       330       340       350       360       370

     210       220       230       240       250       260       270
    NYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCNCCNWSNTGYYT
     || || ||||| || ||||| ||||| |||||||| || |||||||| |||||||| || || || |||||
    CTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTCACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCT
380       390       400       410       420       430       440       450

280       290       300       310       320       330       340       350
NMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGCNYTNAARCCNTG
 || ||||| ||||| ||||| || || || || |||||||| || |||||||| || || || ||||| ||
TCGCTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAGCCCTG
      460       470       480       490       500       510       520

        360       370       380       390       400       410       420
GATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWSNWSNACNYTNCCNCCNCC
||| || || |||||||||| || ||||| ||||| |||||||||| ||||| || || || || || || ||
GATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTGTCAGCCCGACTCCTCAACCCTGCCACCCCC
    530       540       550       560       570       580       590

     430       440       450       460    X
NTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCN
 ||||| || || || || ||||| || || || || || || ||
ATGGAGTCCCCGGCCCCTGGAGGCCACAGCCCCGACAGCCCCGCAGCCCCCTC
  600       610       620       630       640
```

3. US-08-162-407-6 (1-465)
   MUSLIGAND Mouse flt3 ligand mRNA, complete cds.

```
 LOCUS       MUSLIGAND      829 bp ss-mRNA            ROD       04-JAN-1994
 DEFINITION  Mouse flt3 ligand mRNA, complete cds.
 ACCESSION   L23636
 KEYWORDS    ligand.
 SOURCE      Mus musculus (strain SJL/J) cDNA to mRNA.
   ORGANISM  Mus musculus
             Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
             Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
 REFERENCE   1  (bases 1 to 829)
   AUTHORS   Lyman,S.D., James,L., VandenBos,T., deVries,P., Brasel,K.,
             Gliniak,B., Hollingsworth,L., Picha,K.S., McKenna,H.J.,
             Splett,R.R., Fletcher,F.A., Maraskovsky,E., Farrah,T.,
             Foxworthe,D., Williams,D.E. and Beckmann,M.P.
   TITLE     Molecular cloning of a ligand for the flt3/flk-2 tyrosine kinase
             receptor -- a proliferative factor for primitive hematopoietic
             cells
   JOURNAL   Cell 75, 1157-1167 (1993)
   STANDARD  full automatic
 COMMENT     NCBI gi: 439441
```

```
FEATURES             Location/Qualifiers
     source          1..829
                     /organism="Mus musculus"
                     /strain="SJL/J"
                     /sequenced_mol="cDNA to mRNA"
     sig_peptide     32..109
                     /codon_start=1
     CDS             32..727
                     /note="NCBI gi: 439442"
                     /codon_start=1
                     /product="flt3 ligand"
                     /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                     KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                     NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLALKPCIGKACQNFSRCLEVQ
                     CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLPLTLVLLAAAWGLRWQRARRR
                     GELHPGVPLPSHP"
     mat_peptide     110..724
                     /codon_start=1
BASE COUNT      163 a    265 c    221 g    180 t
ORIGIN

Initial Score    =    345  Optimized Score  =    400  Significance = 23.16
Residue Identity =     46%  Matches          =    322  Mismatches   =   143
Gaps             =      8  Conservative Substitutions               =     0

             X       10        20        30        40        50        60
             CARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARATNMGNGARYT
             |  |||||||  |||   ||||| || || || ||  |||||   || ||| | || |||||
   GCGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAGTTTAGAGAGTT
     110     X 120       130       140       150       160       170

         70        80        90       100       110       120       130
   NWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYTNTGYGGNGG
    || ||| |||| ||  ||||||||||| || || || ||    ||||| |||||||| ||| |||   |
   GACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAGAAGCACTGCAAGGC
   180       190       200       210       220       230       240

      140       150       160       170       180       190       200
   NYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNAARATGCARGG
    || ||||| ||  | || || ||||| ||||| ||||  || ||||| || || || || ||||||||||
   CTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATGCAAAC
 250       260       270       280       290       300       310       320

    210       220       230       240       250       260       270
   NYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCNCCNWSNTGYYT
    || ||  |||   || ||||| ||||| |||||||| ||   |||| | |||||||| |  ||    |||||
   GCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCTACCAGAATGTCT
       330       340       350       360       370       380       390

 280       290       300       310       320       330       340
   NMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGCNYTNAARCCNT-
    || ||||| ||||| ||||| || |  || ||  |||| || ||   ||||||  | || || ||||| |
   GCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCTGCTTGCTCTGAAGCCCTG
      400       410       420       430       440       450       460

    350       360       370       380       390       400       410
   -----GGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWSNWSNACNYTNCC
        |||   |  | |||||||||| || |||| ||| | |||||||||| ||||| || || || |
   TATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCT
     470       480       490       500       510       520       530
```



```
      420       430       440       450       460    X
   NCCNCCNTGGWSNCCNMGNCCNYTNGARGCNAC--NGCNCCNACNGCNCCN
    || ||  |||| || |   | || ||||| ||   || || |  ||  |
   GCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCAG
   540       550       560       570       580       590
```

4. US-08-162-407-6 (1-465)
   MMU04807 Mus musculus FLT3/FLK2 ligand mRNA, complete cds.

```
LOCUS       MMU04807     1152 bp    mRNA            ROD       11-MAY-1994
DEFINITION  Mus musculus FLT3/FLK2 ligand mRNA, complete cds.
ACCESSION   U04807
KEYWORDS    .
SOURCE      mouse.
  ORGANISM  Mus musculus
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Rodentia; Myomorpha; Muridae; Mus.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Hannum,C., Culpepper,J., Campbell,D., McClanahan,T., Zurawski,S.,
            Bazan,J.F., Kastelein,R., Hudak,S., Wagner,J., Mattson,J., Luh,J.,
            Duda,G., Martina,N., Peterson,D., Menon,S., Shanafelt,A.,
            Muench,M., Kelner,G., Namikawa,R., Rennick,D., Roncarolo,M.G.,
            Zlotnik,A., Rosnet,O., Dubreuil,P., Birnbaum,D. and Lee,F.
  TITLE     Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of
            haematopoietic stem cells and is encoded by variant RNAs
  JOURNAL   Nature 368, 643-648 (1994)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1152)
  AUTHORS   Culpepper,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1993) Janice A. Culpepper, DNAX Research
            Institute, 901 California Avenue, Palo Alto, CA 94304, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 483846
FEATURES             Location/Qualifiers
     source          1..1152
                     /clone="mouse Flt3/Flk2 ligand T110"
                     /clone_lib="TA4 cDNA"
                     /organism="Mus musculus"
                     /cell_line="TA4 stroma"
     CDS             256..954
                     /note="NCBI gi: 483847"
                     /codon_start=1
                     /product="FLT3/FLK2 ligand"
                     /translation="MTVLAPAWSPNSSLLLLLLLLSPCLRGTPDCYFSHSPISSNFKV
                     KFRELTDHLLKDYPVTVAVNLQDEKHCKALWSLFLAQRWIEQLKTVAGSKMQTLLEDV
                     NTEIHFVTSCTFQPLPECLRFVQTNISHLLKDTCTQLLGLKPCIGKACQNFSRCLEVQ
                     CQPDSSTLLPPRSPIALEATELPEPRPRQLLLLLLLLPLTLVLLAAAWGLRWQRARR
                     RGELHPGVPLPSHP"
BASE COUNT      238 a    346 c    324 g    244 t
ORIGIN

Initial Score    =    343  Optimized Score  =    399  Significance = 23.01
Residue Identity =     46%  Matches          =    321  Mismatches   =   144
Gaps             =      8  Conservative Substitutions               =     0

             X       10        20        30        40        50        60
             CARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARATNMGNGARYT
```

```
                 |  |||||||  |||   |||||  ||  ||  ||  ||  |||||   ||  |||  |  ||  |||||
    GCGGGGGACACCTGACTGTTACTTCAGCCACAGTCCCATCTCCTCCAACTTCAAAGTGAAGTTTAGAGAGTT
330         340         350         360         370         380         390         400

          70          80          90         100         110         120         130
    NWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYTNTGYGGNGG
     ||  |||  ||||  ||  ||||||||||  ||  ||  ||  ||     |||||  |||||||||  |||  |||    |
    GACTGACCACCTGCTTAAAGATTACCCAGTCACTGTGGCCGTCAATCTTCAGGACGAGAAGCACTGCAAGGC
          410         420         430         440         450         460         470

       140         150         160         170         180         190         200
    NYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNAARATGCARGG
     ||  |||||  ||   |  ||  ||  |||||  |||||  ||||   ||  |||||  ||  ||  ||  ||  ||||||||||
    CTTGTGGAGCCTCTTCCTAGCCCAGCGCTGGATAGAGCAACTGAAGACTGTGGCAGGGTCTAAGATGCAAAC
        480         490         500         510         520         530         540

     210         220         230         240         250         260         270
    NYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCNCCNWSNTGYYT
     ||  ||  |||    ||  |||||  |||||  ||||||||  ||    ||||  |  ||||||||  |  ||     |||||
    GCTTCTGGAGGACGTCAACACCGAGATACATTTTGTCACCTCATGTACCTTCCAGCCCCTACCAGAATGTCT
      550         560         570         580         590         600         610

   280         290         300         310         320         330         340
    NMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGTNGCNYTNAARCCNT-
     ||  |||||  |||||  |||||  ||  |  ||  ||  ||||  ||  ||    ||||||  |  |  ||  |||||  |
    GCGATTCGTCCAGACCAACATCTCCCACCTCCTGAAGGACACCTGCACACAGCTGCTTGGTCTGAAGCCCTG
    620         630         640         650         660         670         680

       350         360         370         380         390         400         410
    -----GGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWSNWSNACNYTNCC
          |||    |  |  ||||||||||  ||  |||||  |||  |  ||||||||||  |||||  ||  ||  ||  |
    TATCGGGAAGGCCTGCCAGAATTTCTCTCGGTGCCTGGAGGTGCAGTGCCAGCCGGACTCCTCCACCCTGCT
690         700         710         720         730         740         750         760

      420         430         440         450         460    X
    NCCNCCNTGGWSNCCNMGNCCNYTNGARGCNAC--NGCNCCNACNGCNCCN
     ||  ||  ||||  ||  |    |  ||  |||||  ||    ||  ||  |   ||   |
    GCCCCCAAGGAGTCCCATAGCCCTAGAAGCCACGGAGCTCCCAGAGCCTCGGCCCAGGCAG
          770         780         790         800         810 X     820
```

5. US-08-162-407-6 (1-465)
   MZEPOD Maize pyruvate,orthophosphate dikinase mRNA, compl

```
LOCUS       MZEPOD       3171 bp ss-mRNA            PLN       30-SEP-1988
DEFINITION  Maize pyruvate,orthophosphate dikinase mRNA, complete cds.
ACCESSION   J03901
KEYWORDS    pyruvate orthophosphate dikinase.
SOURCE      Maize (strain Golden Cross Bantam) green leaf, cDNA to mRNA, clones
            pPPD[71,1067].
  ORGANISM  Zea mays
            Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
            Commelinidae; Cyperales; Poaceae.
REFERENCE   1  (bases 1 to 3171)
  AUTHORS   Matsuoka,M., Ozeki,Y., Yamamoto,N., Hirano,H., Kano-Murakami,Y. and
            Tanaka,Y.
  TITLE     Primary structure of maize pyruvate,orthophosphate dikinase as
            deduced from cDNA sequence
  JOURNAL   J. Biol. Chem. 263, 11080-11083 (1988)
  STANDARD  full automatic
```



```
COMMENT     Draft entry and printed copy of sequence for [1] kindly provided by
            M.Matsuoka, 18-MAY-1988.

            NCBI gi: 168579
FEATURES             Location/Qualifiers
     source          1..3171
                     /organism="Zea mays"
     mRNA            <1..3171
                     /note="PODK mRNA"
     CDS             114..2957
                     /note="pyruvate,orthophosphate dikinase (EC 2.7.9.1);
                     NCBI gi: 168580"
                     /codon_start=1
                     /translation="MAASVSRAICVQKPGSKCTRDREATSFARRSVAAPRPPHAKARR
                     RHPLRLRRGTGPHCSPLRAVVDAAPIQTTKKRVFHFGKGKSEGNKTMKELLGGKGANL
                     AEMASIGLSVPPGFTVSTEACQQYQDAGCALPAGLWAEIVDGLQWVEEYMGATLGDPQ
                     RPLLLSVRSGAAVSMPGMMDTVLNLGLNDEVAAGLAAKSGERFAYDSFRRFLDMFGNV
                     VMDIPRSLFEEKLEHMKESKGLKNDTDLTASDLKELVGQYKEVYLSAKGEPFPSDPKK
                     QLELAVLAVFNSWESPRAKKYRSINQITGLRGTAVNVQCMVFGNMGNTSGTGVLFTRN
                     PNTGEKKLYGEFLVNAQGEDVVAGIRTPEDLDAMKNLMPQAYDELVENCNILESHYKE
                     MQDIEFTVQENRLWMLQCRTGKRTGKSAVKIAVDMVNEGLVEPRSAIKMVEPGHLDQL
                     LHPQFENPSAYKDQVIATGLPASPGAAVGQVVFTAEDAEAWHSQGKAAILVRAETSPE
                     DVGGMHAAVGILTERGGMTSHAAVVARWWGKCCVSGCSGIRVNDAEKLVTIGSHVLRE
                     GEWLSLNGSTGEVILGKQPLSPPALSGDLGTFMAWVDDVRKLKVLANADTPDDALTAR
                     NNGAQGIGLCRTEHMFFASDERIKAVRQMIMAPTLELRQQALDRLLTYQRSDFEGIFR
                     AMDGLPVTIRLLDHPSYEFLPEGNIEDIVSELCAETGANQEDALARIEKLSEVNPMLG
                     FRGCRLGISYPELTEMQARAIFEAAIAMTNQGVQVFPEIMVPLVGTPQELGHQVTLIR
                     QVAEKVFANVGKTIGYKVGTMIEIPRAALVADEIAEQAEFFSFGTNDLTQMTFGYSRD
                     DVGKFIPVHLAQGILQHDPFEVLDQRGVGELVKFATERGRKARPNLKVGICGEHGGEP
                     SSVAFFAKAGLDFVSCSPFRVPIARLAAAQVLV"
BASE COUNT      691 a    852 c    971 g    657 t
ORIGIN      1 bp upstream of EcoRI site.
```

```
Initial Score   =    161  Optimized Score  =    241  Significance =  9.55
Residue Identity =    28%  Matches         =    224  Mismatches   =   222
Gaps            =     46  Conservative Substitutions              =     0

            X           10          20          30          40          50
            CA--RGAYTGY---WSNTTYCARCAYWSNCCNATNWSNWSNGA-YTTYGCNGTNAARATNMG
              ||  ||||   ||  ||||  |||  |  |  ||  ||  |||    |  |  ||
   AGATGGCGAGCATCGGGCTGTCGGTGCCGCCAGGGTTC-ACGGTGTCGACGGAGGCGTGCCAGCAGTACCAG
      430   X   440         450         460         470         480         490

     60          70          80          90         100         110         120
   NGARYTNWSNGAYTAYYTNYTNCARGAYTAY--CCNGTNACNGTNGCNWSNAAYYTNCA--RGAYGARGARY
      ||   | |     ||   |  | ||  ||  |    | |  |  ||  | ||    ||| ||  || ||  | |
   GACGCCGGGTGCGCCCTCCCCGCGGGGCTCTGGGCCGAGATCGT--CGAC-GGCCTGCAGTGGGTGGAGGAG
      500         510         520         530         540         550         560

     130         140         150         160         170         180         190
   TNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNY---TNAARACNGTNGCNGGNW
   |   |||  | |      |  || |     | |||||   | |   ||  |   |     | |  || |
   TACATGGGCGCC------ACCCTGGGCGATCCGCAGCGCCCGCTCCTGCTCTCCGTCCGCTCCGGCGCCGCCG
      570             580         590         600         610         620         630

      200         210         220         230         240         250         260
   SNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTG--YGCNTTYCARCC
       ||||  ||  |  | ||  |  ||   ||    |   | |  ||  || |||||  |||    | | |
   TGTCCATGCCCGGCATGATGGACACGGTGCTCAACCTGGGGCTCAACG--ACGAAGTGGCCGCCGGGCTGGC
         640         650         660         670         680         690         700
```

```
      270       280       290       300       310       320       330
NCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNGT
  |    ||  |    || |||| |  || | |   || || |  || |  || |  |  | |||||  ||
GGCCAAGAGCGGGGAGCGCTTCG--CCTACGACTCCTTC-CGCCGCTTCCTCGACATGTTCG-GCAAC--GT
       710       720       730       740       750       760

   340       350           360       370       380       390
NG-CNYTNA-ARCCNTGGATNAC---NMG-NCARAA-YTTYWSNMGNTGYYTNGARYTNCARTGYC---ARC
 | | |  | | ||   | | ||    ||   |||| ||| || |  ||    ||||   || | |   ||
CGTCATGGACATCCCCCGCTCACTGTTCGAAGAGAAGCTTGAGCACATG-AAGGAATCCAAGGGGCTGAAGA
 770       780       790       800       810       820       830

400       410       420       430       440       450       460    X
CNGAYW-SNWSNACNYTNCCNC-CNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCN
  |||| | ||  | |  || |   ||    ||  | ||    |  | | | |      | ||  | |
ACGACACCGACCTCACGGCCTCTGACCTCAAAGAGCTCGTGGGTCAGTA-CAAGGAGGTCTACCTCTCAGCC
840       850       860       870       880       890       900    X

 AAGGGAGA
910
```

6. US-08-162-407-6 (1-465)
   HHU06941 Human herpesvirus 6B R1 replication origin (oriLyt

```
LOCUS       HHU06941     1651 bp    DNA             VRL       02-MAY-1994
DEFINITION  Human herpesvirus 6B R1 replication origin (oriLyt).
ACCESSION   U06941
KEYWORDS    .
SOURCE      Human herpesvirus 6B.
  ORGANISM  Human herpesvirus 6B
            Vira; Viruses; dsDNA enveloped viruses; Herpesviridae;
            Betaherpesvirinae.
REFERENCE   1  (bases 1 to 1651)
  AUTHORS   Dewhurst,S., Krenitsky,D. and Dykes,C.
  TITLE     Human herpesvirus 6B origin: sequence diversity, requirement for
            two binding sites for origin-binding protein and enhanced
            replication from naturally occurring origin multimers
  JOURNAL   Unpublished
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1651)
  AUTHORS   Dewhurst,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-1994) Stephen Dewhurst, Microbiology and
            Immunology, University of Rochester Medical Center, 575 Elmwood
            Avenue, Rochester, NY 14642, USA
  STANDARD  full automatic
COMMENT     NCBI gi: 476271
FEATURES             Location/Qualifiers
     source          1..1651
                     /clone="RC3 clone"
                     /strain="R1 isolate"
                     /organism="Human herpesvirus 6B"
                     /map="approx. 75-80kb on R1 genome"
                     /note="HHV-6B"
     rep_origin      394..761
     misc_binding    618..640
                     /bound_moiety="OBP"
     misc_binding    674..696
```



```
                     /bound_moiety="OBP"
     repeat_unit     873..1065
                     /note="type 1 repeat (exactly as Z29)"
                     /rpt_type=DIRECT
     repeat_unit     1066..1252
                     /note="type 1 repeat"
                     /rpt_type=DIRECT
BASE COUNT      555 a    264 c    296 g    536 t
ORIGIN
```

```
Initial Score    =    145  Optimized Score  =    240  Significance =  8.36
Residue Identity =    27%  Matches          =    223  Mismatches   =   231
Gaps             =     43  Conservative Substitutions               =     0

            X         10        20        30        40        50
           CAR--GAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARAT--NMGNG
           |||  |||||||  |   |   ||| |   | || ||  ||  | |    |||||   | |
GTAGGCACCTCAAACGATTGCAAGTGAAATTTTTGTCTGGTTTGCTCACATAAT-CGAATAAGATCACCATG
      30  X     40        50        60        70        80        90

60         70        80        90        100       110       120
ARYTNWSNGAY-TAYYTNYT-NCAR--GAYTAYCCNGTNACNGTNGC-NWSNAAYYTNCARGAYGARGARYT
| || || |.| ||||  ||  | |   | | ||  |  |      |  ||    | |  ||||| | | ||
ATCT-TCGGCCATACCAACTAACTGCTCAGTTCCATGAAAAAAAAACACACGGGCATAAAAGACGTA-ATTT
     100       110       120       130       140       150       160

  130       140       150       160       170       180       190
NTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGA--RMGNYTNA-ARACNGTNGCNGGNWS
  ||   |  || | ||  |     || |   | || ||  |  |  | | |  | |||  ||       |
TAGT-TAGAATTGT-GCCTCCAAAATTAG-TGATAGGTGTTTTAACCGTGGTCGAGAAATCGTTTTCTACGG
     170       180       190       200       210       220       230

   200       210       220       230       240       250       260
NAARATGCARGGNYTNY-TNGARMGNGTNAAYAC-NGARA--TNCAYTTYGTNACNAARTGYGCNTTYCARC
  |||| ||    |  |   ||||   | | |||  |||| | | || | | |   |     |  | ||
TCAAATTCATTAACAACGGCGAAA---TAACTACAGGAGAATTTCTCTCCATTAACGATAAAACAGATAAAA
     240       250       260       270       280       290       300

     270       280       290       300       310       320       330
CNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYAT--NWSNMGNYTNYTNCARGA-RACNWSNGARCARY
|  |    || | |||    ||||   ||||  | ||   || || |  |   || | ||  ||  | || |
CATCCAAAAGTTCTTTCTCTTTTGGAAAAACATAGATCCAAGCAG-CAGTAGAAACAGGAGCAGAAACCACC
     310       320       330       340       350       360       370

      340               350       360       370       380       390
TNGTNGCNYTNA--ARCC------NTGGATNACNMGNCARAAYTTYWSNMGNTG---YYTNGARYTNCARTG
   | |  || |  ||||       ||||||       |||||    ||    ||   ||| | | |  |||
GTTTCGTTTTCATCAGCCATCTTTGTGGATTCGATCACAAAAACAAAGATAATGGGTTTTTGTGGTGAAATC
     380       390       400       410       420       430       440

     400       410       420       430       440       450       460
YCARCCNGAY-WSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNG-ARGCNACNGCNCCNACNGCN
|  |      | || |  ||  |  |    |    |||  | |         | || | ||       |
CTTATATATTATGTTTGAC-GTAACATAACACGCGTCATCAAACATAAAAGTAAACCACAAGTTGAAATATA
   450       460       470       480       490       500       510

  X
CCN
||
CCGTTTTCTATAT
```

```
      520


7. US-08-162-407-6 (1-465)
   TAGLUIN1     Wheat Glu-Aly sublocus with Wis-2 insertion 5' end

 LOCUS       TAGLUIN1      794 bp    DNA             PLN       06-JUL-1989
 DEFINITION  Wheat Glu-Aly sublocus with Wis-2 insertion 5' end (0.8 kb Hind III
             fragment).
 ACCESSION   X05995
 KEYWORDS    Glu-Aly gene; glutenin; insertion sequence; retrotransposon;
             Wis-2 insertion sequence.
 SOURCE      wheat.
   ORGANISM  Triticum aestivum
             Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
             Commelinidae; Cyperales; Poaceae.
 REFERENCE   1  (bases 1 to 794)
   AUTHORS   Harberd,N.P., Flavell,R.B. and Thompson,R.D.
   TITLE     Identification of a transposon-like insertion in a Glu-1 allele of
             wheat
   JOURNAL   Mol. Gen. Genet. 209, 326-332 (1987)
   STANDARD  full automatic
 COMMENT     Sequence comparision with Cheyenne
             Glu-Aly sublocus gene has shown that the coding sequence of the
             Chinese Spring sublocus gene is interrupted by a 8.0 kb insertion
             sequence. A DNA duplication and inverted terminal sequences are
             flanking the insertion sequence. See also x05996.
             Data kindly reviewed (18-Apr-1988) by Flavell R.B.

             NCBI gi: 21780
 FEATURES             Location/Qualifiers
      source          1..794
                      /organism="Triticum aestivum"
                      /strain="Chinese spring"
                      /chromosome="chromosome 1a"
      misc_feature    <1..482
                      /note="glutenin homologous seq. (161 AA; includ. stop
                      codon) (1 is 2nd base in codon)"
      misc_feature    478..481
                      /note="DNA duplication"
      intron          483..>794
                      /note="insertion sequence"
      misc_feature    484..493
                      /note="inverted terminal sequence"
 BASE COUNT      290 a    183 c    178 g    139 t      4 others
 ORIGIN

Initial Score    =    134  Optimized Score  =    228  Significance =  7.55
Residue Identity =    28%  Matches          =    224  Mismatches   =   196
Gaps             =     64  Conservative Substitutions               =     0

       10    X   20        30        40          50          60        70
    YTGYWSNTTYCARC-AYWSNCCNATNWSNWSNGAYTTYGCNGTNA--AR-ATNMGNG-ARYTNWSNGAYTAY
                |||  |||  | |   | ||  |  | ||     |  || |  || | |||  ||  |||||
              AAGCTTCTCCACAACAGCCAGGAAAATGGCAAGAACTAGGACAAGGGCAACAATGGTACTAT
              X       10        20        30        40          50        60

          80        90       100       110       120       130       140
    YTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNA-AYYTNCARGAYGARGARYTNTGYGGNGGNYTNTG--G
    |      ||  ||| | | |   | |  || || | |    ||| | | || |||   ||      || ||  |
```



```
    C-----CA--ACTTCTCTG---CAGCAGCCAGGACAAGGGCAACAAG-GGTACT-ACCGAACTTCTCTGCAG
                  70          80          90         100         110       120

           150       160         170       180       190       200
    MGNYTNGTNYTNGCNCARMGNTGG----ATGGARMGNYTNAARACNGTNGCNGGNWSNAARATGCARGGNYT
    |  |   |   |     ||||   ||    |  ||||  ||   | | |  ||  | || |||  ||||
    CAGCCAGGACAAAGGCAACAAGGGTACTACCGAAC--TTCTCGGCAGCAGCCAGGAC-AAG-GGCAA--CAG
         130       140         150       160       170       180

  210       220       230       240       250       260       270
    NYTNGARMGNGTNAAY-ACNGARATNCAYTTYGTNACNAARTGYGCNTTYCA-RCCNCCNCCNWSNTGYYTN
     |  || |  |  ||| |  | | | | |  | |  | | | || |   ||| | |     | ||   ||
    ATAGGACAATGGCAACAAGGGTACTACCCAACTTCTCCACA-GCAC---CCAGGACAAGGGCAAC-AACCAG
  190       200       210       220       230       240       250

  280       290       300       310       320       330       340
    MGNTTYGTNCARACNAAYATNWSNM-GNYTNYTNCARGARACNWSNGARCARYTNG-----TNGCNYTNAA-
     |    || ||||  ||||  || | | |  | | |||||   |  |||||||  |       ||    ||
    -GACAAGTGCAAA-AAATAGGACAAGGGCAACAACCAGAAAAAGGGCAACAACTAGGACAAGAGCAACAAAT
       260       270       280       290       300       310       320

         350       360       370       380       390         400
    --RCCNTGGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCARTGY--CARCC-NGAYW----SNWS
      | |  ||   || |  ||||||    | |    ||   ||    || ||  |||||  ||||    | ||
    AGGACAAGGGCAAC-AACCAGAACAAGGGCAACAACC---AGGACAAGGGCAACAACCAGGACAAGGGCAAC
      330       340       350       360         370       380       390

    410       420       430       440         450       460    X
    NACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGA----RGCNACNGCNCCNACNGCNCCN
     |   | |  ||    |  ||  | ||    |  ||    ||| ||  |    ||     |
    AAGGGTACTACCCAACTTCTCTGC-AGCAGCCAGGACAAGGGCAACAACCAGGACAATGACAACAACCAGGA
       400       410       420       430       440       450   X   460
```

```
8. US-08-162-407-6 (1-465)
   COTMYBA      Cotton DNA-binding domain mRNA.

 LOCUS       COTMYBA      1006 bp ss-mRNA            PLN       31-DEC-1993
 DEFINITION  Cotton DNA-binding domain mRNA.
 ACCESSION   L04497
 KEYWORDS    .
 SOURCE      Gossypium hirsutum (cultivar Acala SJ-2) 3-day pre-anthesis ovule
             cDNA to mRNA.
   ORGANISM  Gossypium hirsutum
             Eukaryota; Plantae; Embryobionta; Magnoliophyta; Magnoliopsida;
             Dilleniidae; Malvales; Malvaceae.
 REFERENCE   1  (bases 1 to 1006)
   AUTHORS   Wilkins,T.A. and Lu,C.-C.
   JOURNAL   Unpublished (1993)
   STANDARD  full automatic
 COMMENT     NCBI gi: 437326
 FEATURES             Location/Qualifiers
      source          1..1006
                      /organism="Gossypium hirsutum"
                      /cultivar="Acala SJ-2"
                      /dev_stage="3-day pre-anthesis"
                      /sequenced_mol="cDNA to mRNA"
                      /tissue_type="ovule"
      mRNA            1..1006
      CDS             59..943
```

```
                     /note="MYB A; putative;  NCBI gi: 437327"
                     /codon_start=1
                     /translation="MGRSPCCEKAHTNKGAWTKEEDQRLINYIRVHGEGCWRSLPKAA
                     GLLRCGKSCRLRWINYLRPDLKRGNFTEEEDELIIKLHSLLGNKWSLIAGRLPGRTDN
                     EIKNYWNTHIKRKLISRGIDPQTHRPLNQTANTNTVTAPTELDFRNSPTSVSKSSSIK
                     NPSLDFNYNEFQFKSNTDSLEEPNCTASSGMTTDEEQQEQLHKKQQYGPSNGQDINLE
                     LSIGIVSADSSRVSNANSAESKPKVDNNNFQFLEQAMVAKAVCLCWQLGFGTSEICRN
                     CQNSNSNGFYSYCRPLDS"
     protein_bind    92..403
                     /note="putative"
                     /bound_moiety="MYB"
     repeat_region   107..133
                     /note="MYB DNA-binding domain repeat signature 1;
                     putative"
BASE COUNT      323 a    209 c    204 g    270 t
ORIGIN

Initial Score    =    134  Optimized Score  =    249  Significance =  7.55
Residue Identity =    27%  Matches          =    231  Mismatches   =   217
Gaps             =     49  Conservative Substitutions               =     0

            X       10          20          30          40          50
            CARGAYTGYW--SNTTYCARCAYWSNCC-NATNWSNWSNG--AYTTYGCNGT--NAARATNM
             ||| |||||  | ||||    ||| ||  |  || || |   ||| |  ||   ||||
   TGTCCATGGTGAAGGCTGCTGGCGTTCC----CTCCCCAAAGCTGCTGGGCTGCTTAGATGTGGTAAGAGTT
   150       160       170         180       190       200       210

      60          70        80          90        100       110       120
   GNGARYTNWS--NGAYTAYYTNYTNCARGAYT-AYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARY
   |   ||| ||   ||| | || ||  ||| || |||     | |     |  | |    |||| || ||||
   GCAGATTAAGATGGATAAACTACTTGAGGCCTGATCTTAAGAGAGGAAATTTCACTGAAGAAGAAGATGAGC
     220       230       240       250       260       270       280

      130       140       150       160         170       180       190
   TNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGG--ATGGARMGNYTNAARACNGTNGCNGGNWS
   |   |       ||    || ||  | |     ||||  |||  || ||  | |  ||||   |  | || ||
   TTATCATCAAGCTTCACAGTTTACT-TGGAAACAAA--TGGTCATTGATTGCTGGAAGA--TTACCAGGAAG
   290       300       310       320         330       340       350

       200             210       220       230       240       250
   NA-ARAT-------GCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTY
     | ||||         |||  || || ||||      |||   |||   |  || | |  | | |      |
   AACAGATAATGAGATAAAGAACTACTGGAAC--ACACACATCAAAAGAAAGCTTATAAGCAGAGGAATTGAT
      360       370       380         390       400       410       420

  260         270       280       290         300       310         320
   CARCCNCCNC--CNWSNTGYYTNMGNTTYGTNCAR-ACNAAYA-TNWSNMGNYTNYTNCARGA---RACNWS
   | ||   | |  | || | ||       ||  ||| || ||||   || |  |  |  | | |   ||  ||
   CCACAAACTCATCGTCCTCTCAATCAAACGGCCAATACCAACACAGTCACAGCCCCCACCGAATTGGATTTC
      430       440       450       460       470       480       490

       330       340           350       360       370       380       390
   NGARCARYTNGTNGCNYTNAAR---CC--NTGGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCAR
     || |||| |   | |    |   ||   |  |   | |   |||||   || |   | ||  |||| |||
   --AGAAACTCG-CCCACATCCGTTTCCAAATCCAGTTCCATCAAAAACCCGTCTC-TGGATTTCAATTACAA
      500       510       520       530       540       550       560

         400       410       420       430       440       450       460
   TGYCARCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCN
   ||  |     |||| ||  |    |     || | ||      |   |  |  || |  | |   | |   |
```



```
   TG-AATTTCAATTCAAGTCCAACACAGATTCCCTTGAAGAACCCAACTGTACAGCCAGCAGTGGC-ATGACT
      570       580       590       600       610       620       630

     X
   CCN
    |
   ACAGATGAAGAGC
      640
```

9. US-08-162-407-6 (1-465)
   SAUNKORFB S.ambofaciens ORF, complete CDS.

```
LOCUS       SAUNKORFB    1449 bp    DNA             BCT       02-NOV-1993
DEFINITION  S.ambofaciens ORF, complete CDS.
ACCESSION   Z22675
KEYWORDS    .
SOURCE      Streptomyces ambofaciens.
  ORGANISM  Streptomyces ambofaciens
            Prokaryota; Bacteria; Firmicutes; Streptomycetaceae.
REFERENCE   1  (bases 1 to 1449)
  AUTHORS   Aubert,M., Weber,E., Schneider,D., Simonet,J.M. and Decaris,B.
  TITLE     Primary structure analysis of a duplicated region in the
            amplifiable AUD6 locus of Streptomyces ambofaciens DSM40697
  JOURNAL   FEMS Microbiol. Lett. 113, 49-56 (1993)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 1449)
  AUTHORS   AUBERT,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1993) to the EMBL/GenBank/DDBJ databases. Martine
            AUBERT, Laboratoire Genetique et Microbiologie, Universite Nancy I
            Faculte des Sciences, Boulevard des Aiguillettes,
            Vandoeuvre-ls-Nancy, 54506, France
  STANDARD  full automatic
COMMENT     NCBI gi: 297826
FEATURES             Location/Qualifiers
     source          1..1449
                     /organism="Streptomyces ambofaciens"
                     /strain="DSM40697"
     CDS             375..1373
                     /note="NCBI gi: 581595"
                     /codon_start=1
                     /translation="MTRRLAQVAKKVGVSEATVSRVLNDKPGVSEATRQSVLSALDVL
                     GYERPTQLRGERARLVGLVLPELQNPIFPAFAEVIGGALAQQGLTPVLCTQTKGGVSE
                     ADYVELLLQQQVSGVVFAGGLFAQADAPHDHYRLLAERNIPVVLINAAIADLDFPCIA
                     CDDAVAVEQSWRHLTSLGHERIGLVLGPGDHLPSRRKLAAVRAAGGSLGDDLVERSMF
                     SLEGGQAAASRLLERGVTGIICASDPLALGAVRAARRRGLHVPRDVSVVGYDDSAFMT
                     CTEPPLTTVRQPIEAMGRAAVDLLWAQIQGTEVPHSELLFEPELVVRGSTAQASAK"
BASE COUNT      233 a    541 c    454 g    221 t
ORIGIN

Initial Score    =    134  Optimized Score  =    242  Significance =  7.55
Residue Identity =    27%  Matches          =    225  Mismatches   =   215
Gaps             =     44  Conservative Substitutions               =     0

            X       10        20        30        40        50        60
            CARGAYTGYWSNTTYCARCA-YWSNCCNATNWSNWSNGAYTTYGCNGTNAARATNMGNGARY
            | || ||||||  ||||||| |||         | || | |||||| |    | |    | |
   CCGACTACGTCGAG-CTGCTG-CTCCAGCAGCAGGTGTCGGGCGTC-GTCTTCGCCGGCGGACTGTTCGCCC
       690 X      700       710       720       730       740       750
```

```
        70         80         90        100        110        120        130
TNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYTNTGYGGNG
   | |||  ||  |   |  |||||||   |    |  |   | ||| | | |  |    | |   ||  |
AGGCGGACGCCCCCCACGACCACTACCGGCTGCTCGCCGAACGCAACATCCCCGTGGTCCTGATCAACGCCG
        760        770        780        790        800        810        820

    140        150        160        170        180        190        200
GNYTNTGGMG--NYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNA-ARATGC
  |  | |||   || |  ||  | | ||   |  ||  |||    | | | || || |  || |  | | |
--CCATCGCGGACCTCG-ACTTCCCCTGCATCGCCTGCGACG----ACGCC-GTCGCCG--TCGAGCAGTCC
        830        840        850        860            870        880

       210        220        230        240        250           260
ARG-GNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTY----CARCCNCC-NC
 || | |  ||         |   |  |||   | | |||   |   || | |   |    | ||| ||  |
TGGCGCCACCTCACCTCCCTGGGCCACGAG---CGCATCG--GCCTGGTCCTCGGCCCCGGCGACCACCTCC
   890        900        910           920        930        940        950

270        280        290        300        310        320        330
CNWSNTGYY-TNMGNT-TYGTNCARACNAAYATNWSNMGNYT-NYTNCARGA-RACNWS--NGARCARYTNG
| ||  |||   || |  |   |   |   |   | || ||  ||   ||| ||| ||   |||| ||| |
CCTCCCGCCGGAAGCTGGCCGCCGTGCGGGCCGCGGGAGGCTCACTCGGGGACGACCTCGTGGAAC-GCTCG
       960        970        980        990       1000       1010       1020

   340        350        360        370        380        390        400
TNGCNYTNAARCCNTGGATNACNMGNCA--RAAYTTYWSNMGNTGYYTNGARYTNCARTGYCARCCNGAYWS
  |  ||    || ||||   |  |  |  |  | ||| ||    || ||||  | | ||| ||| |  ||
ATGTTCT----CCCTGGAGGGCGGGCAAGCGGCCGCCTCCCGCCTGCTGGAAC-GCGGCGTC-ACCGGCATC
         1030       1040       1050       1060       1070       1080

   410        420        430        440        450        460     X
NWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNACNG-CNCCN
 ||  | |   |  | || | | |   |    || |  | | |  | ||     | | | ||
ATCTGCGCCAGCGACCCCCTCGCCCTCGGCGCCGTCCGCGCCGCCCGCAGGCGCGGACTCCACGTACCCCGC
1090       1100       1110       1120       1130       1140       1150       1160
```

10. US-08-162-407-6 (1-465)
    TAGLU1AG Wheat gene for HMW-glutenin subunit located on chr

LOCUS TAGLU1AG 2915 bp DNA PLN 06-JUL-1989
DEFINITION Wheat gene for HMW-glutenin subunit located on chromosome 1A.
ACCESSION X03042
KEYWORDS direct repeat; glutenin; inverted repeat; seed storage protein; signal peptide; storage protein.
SOURCE wheat.
ORGANISM Triticum aestivum
Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida; Commelinidae; Cyperales; Poaceae.
REFERENCE 1 (bases 1 to 2915)
AUTHORS Forde,J., Malpica,J.M., Halford,N.G., Shewry,P.R., Anderson,O.D., Greene,F.C. and Miflin,B.J.
TITLE The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1A of wheat (Triticum aestivum L.)
JOURNAL Nucleic Acids Res. 13, 6817-6832 (1985)
STANDARD full automatic
COMMENT Though the premature stop codon may not affect gene expression the author suggests that the gene is a pseudogene, because no small in vivo transcript of chromosome 1A was found. Within the coding

sequence are two multiple repeated amino acid motifs, PGQGQQ and GYYPTSLQQ
Data kindly reviewed (19-MAR-1986) by J. Forde.

NCBI gi: 21770

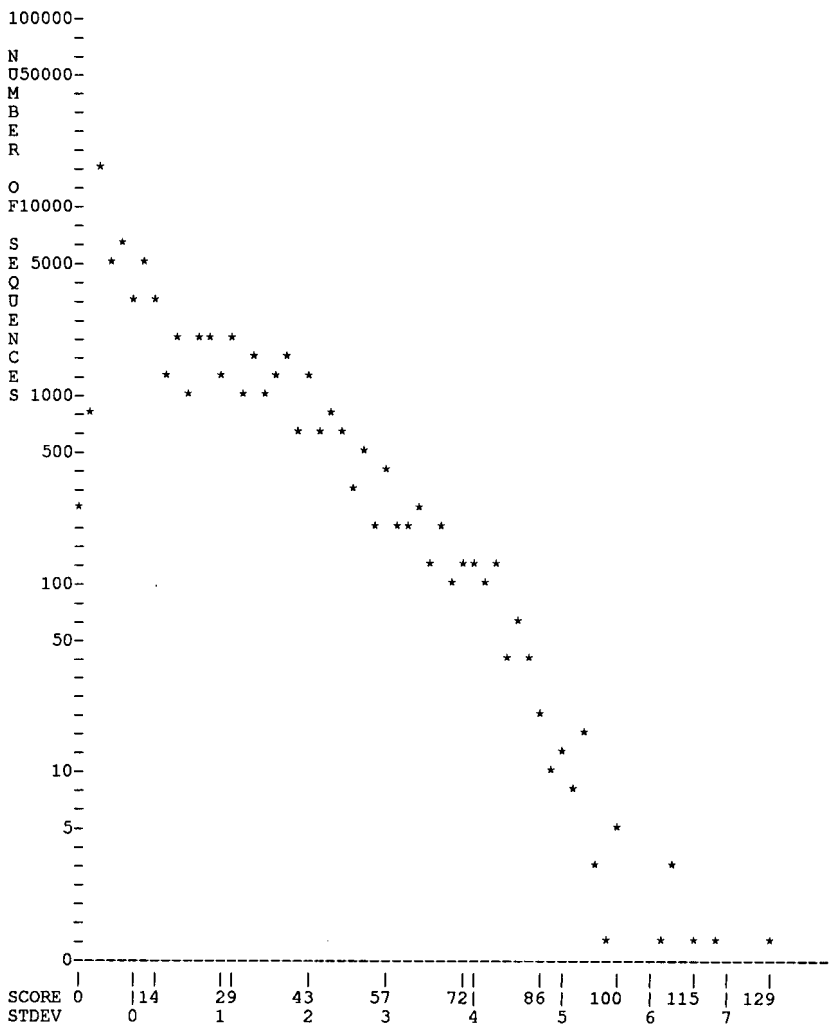
FEATURES	Location/Qualifiers
source	1..2915
	/organism="Triticum aestivum"
repeat_unit	8..24
	/note="inverted repeat X"
repeat_region	243..259
	/note="direct repeat B"
repeat_region	272..291
	/note="imp. direct repeat A"
repeat_region	297..316
	/note="imp. direct repeat A"
repeat_region	318..337
	/note="imp. direct repeat A"
repeat_region	343..359
	/note="direct repeat B"
repeat_unit	419..431
	/note="inverted repeat Y"
promoter	582..586
	/note="pot. CAAT-box"
promoter	620..624
	/note="pot. CAAT-box"
repeat_unit	648..664
	/note="inverted repeat X"
promoter	696..699
	/note="pot. CAAT-box"
promoter	724..730
	/note="put. TATA-box"
CDS	815..2620
	/note="precursor"
	/codon_start=1
CDS	815..877
	/note="put. signal peptide (aa -21 to -1); NCBI gi: 21771"
	/codon_start=1
	/translation="MAKRLVLFATVVIGLVSLTVA"
CDS	878..2620
	/note="mature HMW subunit of glutenin (aa 1-581)"
	/codon_start=1
misc_feature	1832..1834
	/note="premature TAG stop codon"
repeat_unit	2635..2647
	/note="inverted repeat Y"
misc_feature	2678..2683
	/note="put. polyadenylation signal"

BASE COUNT 977 a 736 c 674 g 528 t
ORIGIN

Initial Score = 134 Optimized Score = 241 Significance = 7.55
Residue Identity = 28% Matches = 238 Mismatches = 199
Gaps = 70 Conservative Substitutions = 0

```
          X       10        20        30        40        50        60
          CARGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTY-GCNGTNAARAT-NMGNGAR
          ||| |  | || | | | | ||| || |  |  ||  ||  | ||    |||||  |  |||
AATATCTCCTCAGCAGGGGTCATATTATC-CAGGCCAAGCTTCTC-CACAACAGCCAGGAAAATGGCAAGAA
```

```
    1210      1220      1230      1240      1250      1260      1270

           70        80        90       100       110       120
    YTNWSNGAYTA--YYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYY--TNCARG---AYGARGARYT
    CT-AG-GACAAGGGCAACAATGGTACTATCCAACTTCTCT-GC-AGCAGCCAGGACAAGGGCAACAAGGGTA
     1280      1290      1300      1310      1320      1330      1340

       130       140       150       160       170       180       190
    NT-GYGGNGGNYTNTG--GMGNYTNGTNYTNGCNCARMGNTGG----ATGGARMGNYTNAARACNGTNGCNG
    CTACCGAACTTCTCTGCAGCAGCCAGGACAAAGGCAACAAGGGTACTACCGAAC--TTCTCTGCAGCAGCCA
      1350      1360      1370      1380      1390      1400      1410

              200       210       220       230       240       250
    GNWSNA----ARATGCARGGNYTNYTNGA--RMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNT
    GGACAAGGGCAACAGATAGGACAATGGCAACAAGGGT-ACTACCCAACTTCTCCGCAGCACCCAG-GACAAG
      1420      1430      1440      1450      1460      1470      1480

     260       270       280       290       300       310       320
    TYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNM-GNYTNYTNCARGARACNWSNGAR
    GGCAACAACC----AG--GAC-AAG---TG--CAAA-AAATAGGACAAGGGCAACAACCAGAAAAAGGGCAA
      1490              1500              1510      1520      1530      1540

    330               340               350       360       370       380       390
    CARYTNG-----TNGCNYTNAA---RCCNTGGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCART
    CAACTAGGACAAGAGCAACAAATAGGACAAGGGCAAC-AACCAGAACAAGGGCAACAACC---AGGACAAGG
       1550      1560      1570      1580      1590      1600      1610

             400               410       420       430       440       450
    GY--CARCC-NGAYW-----SNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGA-----RGCNACNGC
    GCAACAACCAGGACAAGGGCAACAAGGGTACTACCCAACTTCTCTGC-AGCAGCCAGGACAAGGGCAACAAC
          1620      1630      1640      1650      1660      1670      1680

         460    X
    NCCNACNGCNCCN
    CAGGACAATGGCAACAACCAGGA
        1690    X 1700
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file translnsq6.res made by maryh on Thu 6 Apr 95 10:02:19-PDT.


Query sequence being compared:US-08-162-407-6 (1-465)
Number of sequences searched:                57621
Number of scores above cutoff:                4579

      Results of the initial comparison of US-08-162-407-6 (1-465) with:
   Data bank : N-GeneSeq 17, all entries
```

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30

Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	18	14	15.49

Times:	CPU	Total Elapsed
	00:02:36.05	00:02:37.00

Number of residues:	24347505
Number of sequences searched:	57621
Number of scores above cutoff:	4579

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 7 standard deviations above mean ****					
1. Q14532	T cell protein 19.1 cDNA (par	1301	129	243	7.16	0
	**** 6 standard deviations above mean ****					
2. Q54029	Flocculation protein coding s	2685	117	249	6.39	0
3. Q45455	Sequence encoding murine bone	2373	114	246	6.20	0
	**** 5 standard deviations above mean ****					
4. Q25812	Clone 45-A.	4300	110	252	5.94	0
5. Q15174	Plasmid pAMVBT4.	4583	109	236	5.87	0
6. N90183	DNA sequence of plant vector	4583	109	236	5.87	0
7. Q63241	Crucifer 1-aminocyclopropane-	5613	108	236	5.81	0
8. N30044	Sequence encoding bovine leuk	1162	100	251	5.29	0
9. Q14235	A.faecalis penicillin acylase	2451	100	246	5.29	0
10. Q03661	Maize C3 sequence encoding ac	2546	100	203	5.29	0
11. Q11176	Clone encoding recombinant hu	4569	100	239	5.29	0
12. N70607	Sequence encoding the melanom	2362	97	240	5.10	0
13. Q55768	Human melanoma-associated ant	2368	96	240	5.03	0
14. N60847	Human pre-prolactin gene.	3621	96	236	5.03	0
	**** 4 standard deviations above mean ****					
15. Q46099	A. rhizogenes NIAES1724 megal	5893	95	239	4.97	0
16. Q49206	Golden hamster Aphrodisin cod	501	93	236	4.84	0
17. Q49205	Field hamster Aphrodisin codi	501	93	236	4.84	0
18. Q49210	Hamster Aphrodisin coding seq	713	93	211	4.84	0
19. Q48773	Aphrodisine.	713	93	211	4.84	0
20. Q49207	Field hamster Aphrodisin codi	747	93	237	4.84	0

1. US-08-162-407-6 (1-465)
Q14532 T cell protein 19.1 cDNA (partial).

ID Q14532 standard; DNA; 1301 BP.
AC Q14532;

DT 29-JAN-1992 (first entry)
DE T cell protein 19.1 cDNA (partial).
KW Multiple membrane spanning protein; T cell; development; ss.
OS Mus musculus.
PN WO9116430-A.
PD 31-OCT-1991.
PF 12-APR-1990; UO2518.
PR 13-APR-1990; US-509684.
PA (REDE-) RES DEV FOUND.
PI Macleod CL;
DR WPI; 91-339818/46.
PT New recombinant polypeptide comprising a T-cell protein - used to
PT regulate T-cell development and tumorigenic phenotype and to
PT block T-cell activation in auto:immune disease
PS Disclosure; Fig 20; 98pp; English.
CC An SL 12.4 cDNA library was constructed from which six cDNAs were
CC isolated via subtractive hybridisation against a related sister
CC lymphoma cell line, SL 12.3. Five clones were obtd. representing
CC genes which are preferentially expressed in the SL 12.4 T cell
CC clone, and are undetectably or weakly expressed in SL 12.3.
CC These clones are 19.5 (Q14534), 20.5 (Q14533), 19.1 (Q14532),
CC 19.2 (Q14531) and 19.4 (Q14530).
CC No predicted protein sequence is given, incomplete cDNA clone,
CC therefore, no long ORF.
CC See also Q14530-34.
SQ Sequence 1301 BP; 380 A; 291 C; 241 G; 389 T;

Initial Score	= 129	Optimized Score	= 243	Significance	= 7.16
Residue Identity	= 26%	Matches	= 224	Mismatches	= 221
Gaps	= 44	Conservative Substitutions			= 0

```
            X          10         20         30         40         50
            CARGA---YTGYWSNTTYCA-RCAYWSNCCNATNWSNWS-NGAYTTYGCNGTNAARATNMGN
                 | |||   ||||||  ||  | || |||     | |  ||   |  || | ||  ||| |
   TCCCTACCAACTGGAAAGTTGTTGTTTAAATACCCTGTTGGTTAAAATGTCAAACTTACTGTTTAAA--CAC
        620 X      630        640        650        660        670        680

    60         70         80         90        100        110        120
   GARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNA-AYY-TNCARGAYGARGARYTNT
     ||| ||   ||  ||  ||  ||   ||| |  |  |  | | |  || | ||| | ||||  |||||
   TCATT-AGCCATATCCAACTTGAAACATATGCTATTGC--TTG--ACCATATTAAGCCAGACTTTGAACTAG
         690        700        710        720        730        740

   130        140        150        160        170        180        190
   GYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNAARA
   ||    | |  ||||  |     |  |  |||| |   | |      |  ||||      | ||  | || |
   GTCAATGTCCCTGGAGTATAAATGTACCATAGAGCTTCCTTGCTTCTTGCAAAAAGTCCTCAGG--CGAACA
 750        760        770        780        790        800        810

 200        210          220        230        240        250        260
   T-GCARGGNYTNYTNGARMG---NGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCN
   |  |   | |  |  ||||     | | | | | | | | | | |  ||   | |||  |||
   TAACTTTGACCCCCATAAAGGTCCCTCACTTCTCACAGCCTTGATTTTGCTGAA--GACCTTCACACCCTGG
 820        830        840        850        860        870        880

   270        280        290        300        310        320        330
   CCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNY-TNCARGARACNWSNGARCA---RYTNGT
   || |    |||  |  ||| |  ||  || | || |  || | |       || ||  ||||   ||  |
   CCCA---ATCTAGGAAATGTTC-TACTCATGTAAGAATTCTTCATCTCTCCCACCTCATAACAAGTGTCCTT
 890          900        910        920        930        940        950
```

```
   340       350       360       370       380       390       400
NG-CNYTNAARCCNTGG-ATNACNMGNCARAAYTTYWSNMGNTGYY---TNGARYTNCARTGYCARCCNGA-
 | | ||  ||   ||| ||    ||  ||||| ||| |    ||   | | ||  |||| |  | ||
TGTCCCTATAATAATGGCAT----AG-AAAAATCCCTGAAAGACCCAAATTGGGCAACAATACAGATCAGAT
 960       970            980       990      1000      1010      1020

       410       420       430       440       450       460    X
YWS-NWSNACNYTNCCNCCNCCNTGGWS--NCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCN
|||   |  | |   | | |  |||||      |   | |  |||     | | |     |
CAGAGCCTTCATGGGCCCGACAATGGAGAATTGTATTTCTTAAAAAGTGTGTGATCATATTTGTCTGAGCTG
   1030      1040      1050      1060      1070      1080      1090

CAATA
 1100
```

2. US-08-162-407-6 (1-465)
   Q54029 Flocculation protein coding sequence of Saccharomy

```
ID   Q54029 standard; DNA; 2685 BP.
AC   Q54029;
DT   19-JUL-1994  (first entry)
DE   Flocculation protein coding sequence of Saccharomyces cerevisiae.
KW   Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
KW   Major cell wall protein; glycosyl-phosphatidyl-inositol;
KW   anchoring protein; alpha factor; alpha-agglutinin; invertase;
KW   inulinase; alpha-amylase; Saccharomyces cerevisiae;
KW   floculation protein; enzymatic process; fermentation;
KW   biodegradation; catalysis; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   CDS             1..2685
FT   /*tag=  a
FT   /product= Flocculation protein.
PN   WO9401567-A.
PD   20-JAN-1994.
PF   07-JUL-1993; E01763.
PR   08-JUL-1992; EP-202080.
PR   14-DEC-1992; EP-203899.
PA   (UNIL ) UNILEVER NV.
PA   (UNIL ) UNILEVER PLC.
PI   Klis FM,  Schreuder MP,  Toschka H,  Verrips CT;
DR   WPI; 94-035071/04.
DR   P-PSDB; R47578.
PT   Immobilisation of enzymes to microbial cell wall - by prodn. of
PT   fusion protein of enzyme linked to anchoring protein
PS   Claim 6; Page 59-64; 99pp; English.
CC   The flocculation protein is used in a method to immobilise enzymes
CC   to a microbial cell wall.  The coding sequence (FLO 1) is used in
CC   the production of a recombinant polynucleotide which comprises a
CC   structural gene encoding a protein with catalytic activity and at
CC   least part of a gene encoding at least the C-terminus of a protein
CC   capable of anchoring in a eukaryotic or prokaryotic cell wall.  The
CC   anchoring fragment or protein is selected from alpha agglutinin,
CC   AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a
CC   proteinase of lactic acid bacteria.  The recombinant polynucleotide
CC   preferably also comprises a sequence encoding a signal peptide to
CC   ensure secretion of the expressed product.  The signal peptide is
CC   preferably derived from glycosyl-phosphatidyl-inositol, anchoring
```



```
CC   protein, alpha factor, alpha-agglutinin, invertase or inulinase,
CC   alpha-amylase of Bacillus or proteinases of lactic acid bacteria.
CC   The host microorganism can be used for performing enzymatic
CC   processes on an industrial scale.
SQ   Sequence  2685 BP;   751 A;   699 C;   492 G;   743 T;
```

```
Initial Score    =    117  Optimized Score  =    249  Significance =  6.39
Residue Identity =    28%  Matches          =    242  Mismatches   =   203
Gaps             =     69  Conservative Substitutions               =     0
```

```
            X          10          20          30          40
            CARGAYT----GYWSNT--TYCA---RCAYWSNCCNAT----NWSNWSNGAYTTYGCNGTNA
            |||  ||    |||| |  | ||   || |||  |  |     || || |    ||| |
   GACGACTCTGCAATTCTATCAGTAGGTGGTGCAACCGCGTTCAACTGTTGTGCTCAACAGCAACCGCCGATC
   480       490       500       510       520       530       540

  50        60        70        80        90        100       110
   ARATNMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGA-YGA-
   | || |  |||| |   |||  |  |  |||| | ||   ||     |  |  ||  |||| || | |||
   ACATCAACGAACTTTACCATTGACGGTATCAAG-CCATGGGGT---GGAAGTTTGCCACCT--AATATCGAA
 550       560       570       580       590           600          610

 120       130       140       150       160       170       180
   RGARYTNTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGC---
   |||||  | |    |  |  |  | ||  | ||  | ||| |  || ||    |  |  ||  | ||  |
   GGAACCGTCTATATG--TACGCTGGCTACTATTATC-CAATGAAGGTTGTTTA-CTCGAACGCTGTTTCTTG
     620       630       640       650       660       670       680

 190           200       210       220       230       240       250
   NGGNWS----NAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNT
    || ||      || |   | | |  ||  |||  || | ||| | ||   ||  | | |  | ||| || | |
   GGGTACACTTCCAATTAGTGTGACACTTCCAGATGGT-ACCACTGTAAGTGATGACTT--CGAAGGGTACGT
       690       700       710       720       730       740       750

    260       270       280       290       300       310       320
   TYCARCCNCCNCCNWSNTG-YYTNMGNTTYGTNCARACNAAYATNWSNM-GNYTNYTNCARGAR-AC-NWSN
    |   ||      |  | || ||| ||  ||   | |||      | || | | |  ||  ||  |  |  ||
   CTATTCCTTTGAC--GATGACCTAAG--TCAATCTAACTGTACTGTCCCTGACCCTTCAAATTATGCTGTCA
        760       770       780       790       800       810       820

      330       340       350             360       370       380
   GARCARYT--NGTNGCNYTNAARCCNTGGA-------TNACNMGNC-ARAAYTTYWSNMGNTGYYTNGARYT
   | || |||      |     |||| ||||         |  | |  | | | || ||| |    ||     |
   GTACCACTACAACTACAACGGAACCATGGACCGGTACTTTCACTTCTACATCTACTGAAATGACCACCGTCA
         830       840       850       860       870       880       890

    390           400       410       420       430       440
   NCARTGYCARC------CNGAYWSNWSNACNYTNCCNCCNCCNT--GGWSNCCNMGNCCNYTNGARG-CNA-
    | || |||||      |  |||| || |  |   |     | |  | || || |  ||  |  ||| | |
   CCGGTACCAACGGCGTTCCAACTG-ACGAAACCGTCATTGTCATCAGAACTCCAA--CCAGTGAAGGTCTAA
        900       910       920       930       940          950       960

   450       460    X
   -CNGCNC--CNACNGCNCCN
    | || |  | || |  ||
   TCAGCACCACCACTGAACCATGGACTGGCA
         970       980       990
```

3. US-08-162-407-6 (1-465)

```
Q45455       Sequence encoding murine bone-related sulphatase 1

ID   Q45455 standard; cDNA to mRNA; 2373 BP.
AC   Q45455;
DT   12-OCT-1994  (first entry)
DE   Sequence encoding murine bone-related sulphatase like  precursor
DE   protein OSF-8.
KW   OSF-8; bone-related sulphatase-like protein; osteoporosis;
KW   Paget's disease; osteomalacia; therapy; diagnosis; ss.
OS   Mus musculus osteoblastic cell line MC3T3-E1.
FH   Key             Location/Qualifiers
FT   CDS             150..1820
FT   /*tag= a
PN   AU9344921-A.
PD   03-MAR-1994.
PF   27-AUG-1993; 044921.
PR   28-AUG-1992; JP-230030.
PR   03-DEC-1992; JP-324034.
PA   (FARH ) HOECHST JAPAN LTD.
PA   (FARH ) HOECHST JAPAN KK.
PI   Amann E,  Ito T,  Otawara-hamamoto Y,  Takeshita S;
DR   WPI; 94-109856/14.
DR   P-PSDB; R51355.
PT   Bone-related sulphatase-like protein, OSF-8 - degrades sulphate
PT   groups of proteoglycan sugar chains, useful for treating bone
PT   metabolic diseases
PS   Claim 3; Page 17-22; 29pp; English.
CC   cDNA of mouse OSF-8 (mOSF-8) was isolated from mouse osteoblastic
CC   cell line MC3T3-E1 cDNA library by a combination of PCR and the
CC   subtraction method, and by the differential screening method. OSF-8
CC   has homology with known representative sulphatase molecules but
CC   belongs to a new subclass different from those so far reported.
CC   The inventors claim the precursor protein (R51355 AAs 1-556) which
CC   includes the signal peptide(AAs 1-18),and they also claim a protein
CC   comprised of residues 19-556. They also claim DNA or RNA encoding
CC   the claimed proteins.
SQ   Sequence   2373 BP;    697 A;    547 C;    466 G;    663 T;

Initial Score    =    114  Optimized Score  =    246  Significance =  6.20
Residue Identity =    29%  Matches         =    236  Mismatches   =   202
Gaps             =     53  Conservative Substitutions              =     0

            X       10        20        30        40        50
            CARGAYTGYWSNTTYCARCAYWS--NCCNATNWSNWSNGAYTTYGC--NGTNAARATNM-GN
             ||||||  || |||||||| |||      |  |  |   ||||| |    | ||  | | |
  TCCTTCGATGGAAGACT-AACATTTCAAC-CAGGAAGTCAGGTAGTAAAACTTCCCTTCATTAACTTCATGA
280       290       300       310       320       330       340

   60        70        80        90       100       110       120
  GARY-TNWSNGAYTA-YYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYTNT
  ||||   ||  || | ||| ||  | | |||| |  | |      | ||   | | ||  ||| | |||  |
  GAGCACATGGCACCACCTTCCTAAATGCCTACAC--TAATTCACCCATCTGCTGTCCATCACGTGCAGCAAT
350       360       370       380       390       400       410

  130       140       150       160       170       180       190
  GYG--GNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGAR-MGNYTNAARACNGTNGCNGGNWSNA
  |||  | || || |  |  ||    || |  |||| |  ||||| |  || |||   ||  | || || |
  GTGGAGTGGCCTCTTCA--CT--CACTTG-ACAGA-ATCT-TGGAATAATTTTAAG--GGT--CTGGATCCA
420       430       440       450       460       470
```



```
    200       210       220       230       240       250       260
  ARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAY-TTYGTNACNAARTGYGCNTTYCARCCNCCN
  || |  |    |  |  | | | |  | ||  |||| |||   | | |   |  |    ||  ||
  AATTATA----CGACATGGATG-GACATCATGGAGAAGCATGGCTATCAGACACAGAAATTTGGAAAAGTGG
480          490       500       510       520       530       540

  270       280       290       300       310       320       330
  CCNWSNTGYYTNMG--NTTYGTNCAR--ACNAA-YATNWSNMGNYTNYTNCARGARACNWSNGARCARYTNG
   | |  | ||| ||    || | | |  |  || | | || |  |     |||||||  ||   || |||
  ACTA--TACTTCAGGACATCATTCCATTAGTAACCGTGTGGAAGCATGGACAAGAGA--TG-TTGC-ATTCT
   550         560       570       580       590       600           610

     340       350       360       370       380       390
  TNGCNYTNAARCCNTGGATNACNMGNCA--RAAYTTYWSNM-GNT--GYYTNGARYTNCARTGYCARCC--N
  | || |  | |    ||   || |   |  || ||| || | | |  |  | |||    ||||   |||
  T-GCTCCGACAAGAAGGCAGACCCATAATTAATCTTATCCCTGATAAGAATAGAAGGAGAGTGATGACCAAG
       620       630       640       650       660       670       680

400       410       420       430       440       450       460    X
  GAYW-SNWSNACNYTNCCNCCNCCNTGG--WSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCN
  |||| | || |  |   |    |  | | || |  || |   |  |    ||    ||    |  |
  GACTGGCAGAATACAGACAAAGCAATCGAATGGCTAAGACAGGTTAACTACACCAAGCCATTTGTCCTTTAC
      690       700       710       720       730       740       750 X

  TTGGGAT
    760


4. US-08-162-407-6 (1-465)
   Q25812       Clone 45-A.

ID   Q25812 standard; cDNA; 4300 BP.
AC   Q25812;
DT   04-JAN-1993  (first entry)
DE   Clone 45-A.
KW   GLU-G-R; Xenopus oocyte; cerebellum sublibrary; voltage clamp assay;
KW   neurotransmitter; glutamate; neuronal excitation; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   CDS             377..3676
FT   /*tag= a
FT   /label= Subtype_1a
PN   WO9210583-A.
PD   25-JUN-1992.
PF   12-DEC-1991; U09422.
PR   12-DEC-1990; US-626806.
PR   30-JAN-1991; US-648481.
PR   18-MAR-1991; US-672007.
PA   (UNIW ) UNIV WASHINGTON.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Almers W, Hagen FS, Houamed KM, Mulvihill ER;
DR   WPI; 92-234638/28.
DR   P-PSDB; R25080.
PT   Mammalian G protein-coupled glutamate receptors - activate
PT   phospholipase C or stimulate inositol phospholipid metabolism,
PT   for use in diagnosis and identification of receptor agonists and
PT   antagonists
PS   Disclosure; Fig 5; 144pp; English.
CC   The sequence given encodes mammalian G protien-coupled glutamate
CC   receptor (GLU-G-R) subtype 1a.  The clone was isolated from Xenopus
```

```
CC   oocyte cultures which had been injected with RNA extracted from a
CC   cerebellum sublibrary.  Expression of GLU-G-R activity was assessed by
CC   voltage clamp assay.  Plasmid DNA encoding GLU-G-R activity was
CC   isolated by replica-plating experiments.  The protein encoded by clone
CC   45-A was found to act as a receptor for the neurotransmitter glutamate,
CC   and may play a critical role in the control of neurons.  GLU-G-R and
CC   antibodies may be used to identify agonists and antagonists of
CC   GLU-G-R-mediated neuronal excitation and in diagnosis.
SQ   Sequence   4300 BP;    1038 A;    1191 C;    1148 G;    923 T;

Initial Score    =    110  Optimized Score  =    252  Significance =  5.94
Residue Identity =    29%  Matches          =    241  Mismatches   =   196
Gaps             =     56  Conservative Substitutions              =     0

            X       10        20        30        40        50
            CARGAYTGYWSNTTYCARCAY--WSNCCNATNWSNWSNGAYTTYGCNGTNAA--RATNMGNG
            ||| ||||| | | || || |  || ||    || ||  ||||| | |  ||  ||  |  |
   TCGGTGATGGCAA-ACTGC-CGTGCC-GCTCCAACACCTTCCTCAAC--ATTTTCCGGAGAAAGAAGCCCGG
     2960    X 2970        2980      2990        3000      3010      3020

   60        70        80        90        100         110       120
   ARYTNWSNGAYTAYYTNYTNCAR-GAYTAYCCNGTNACNGTNGCNWSNAAYY---TNCARGAYGARGARYTN
    ||  || || | ||  ||  || | | |  | ||  |  | |  || ||||   |  ||||| ||| ||
   GGC--AGGGAATGCCAATTCTAACGGCAAGTCTGTGTC-ATGGTCTG-AACCAGGTGGAAGAC-AGGCGCCC
          3030      3040      3050       3060       3070      3080

    130         140       150       160       170       180       190
   TGYGGNGGNY---TNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSN
      ||    |   | ||||  |   | | |  | ||| || |  || |  |   ||||  |   | |  ||
   AAGGGACAGCACGTGTGGCAGCGCCTCTCTGTGC-ACG-TGAA--GACCAAC--GAGACGG--CCTGTAACC
 3090      3100      3110      3120          3130        3140        3150

     200       210       220       230       240       250       260
   AARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNC--AYTTYGTNACNA-ARTGYG--CNTTYCA--
   |||  || |    |  |  |||   | | ||   || | |  |   | |  | | ||  ||  | |||||
   AAACAGCCG----TAATCAAACCCCTCACTA-AAAGTTACCAAGGCTCTGGCAACAGCCTGACCTTTTCAGA
       3160      3170      3180      3190      3200      3210

          270       280       290       300       310       320       330
   -RCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAYATNWSNMGNYTNYTNCARGARACNWSNGARCARY
    |||  | || ||   ||| |    |||   ||  || | ||  | |  || | |   || || |  |  |
   TGCCAGCACCAAGACCCTTTA-CAATGT--GGAAGAAGAGGACAATACCCCTTCTGCTCACTTCAGCCCTCC
 3220      3230      3240          3250      3260      3270      3280

        340       350       360       370       380       390
   TNGTNGCNYTNAARCCNTGGATNACNMGNCARAAYTTYWSNMGNTGYYTNGA-RY-TNCARTGYCA-----RC
     |  || ||   |   |||   || || | |     ||| ||  ||| . | ||   |  |||||    |
   CAGCAGCCCTTCTATGGTGGTGCAC-CGAC-GCGGGCCACCCGTGGCCACCACACCACCTCTGCCACCCCAT
 3290      3300      3310         3320      3330      3340      3350

   400       410         420       430       440       450       460    X
   CNGAYWSNWSNACNYTNCCNCC--NCCNTGG-WSNCCNMGNCCNYTNGARGCNACNGCNCCNACNGCNCCN
   | ||| | || |     || ||    | ||| ||   || | |  |||     ||  |  |  || ||  |
   CTGACCGCAGAAGAGACCCCCCTGTTCCTGGCTGATTCCGTCATCCCCAAGGGCTTGCCTCCTCCTCTCCCG
 3360      3370      3380      3390      3400      3410      3420      3430

   CAGCAGCAG
```



```
5. US-08-162-407-6 (1-465)
   Q15174       Plasmid pAMVBT4.

ID   Q15174 standard; DNA; 4583 BP.
AC   Q15174;
DT   11-MAR-1992  (first entry)
DE   Plasmid pAMVBT4.
KW   Bacillus thuringiensis; transgenic; tree; poplar; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   CDS             482..2427
FT   /*tag=  a
FT   misc_feature    480..897
FT   /*tag=  b
FT   /note= "altered synthetic sequence "
PN   WO9118094-A.
PD   28-NOV-1991.
PF   23-MAY-1991; U03629.
PR   24-MAY-1990; US-530133.
PA   (AGRA-) AGRACETUS INC.
PA   (WISC ) WISCONSIN ALU RES FOUND.
PI   McCabe DE, McCown BH.
DR   WPI; 91-369248/50.
DR   P-PSDB; R15356.
PT   Particle-mediated transformation of woody plants - for
PT   construction of trangenics(s) expressing high levels of Bacillus
PT   toxin
PS   Example; Fig 6,7; 43pp; English.
CC   The sequence is that of plasmid pAMVBT4 which encodes a wild type
CC   Bacillus thuringiensis (Bt) gene from Bt strain HD-1-Dipel. This
CC   codes for an insect toxin amino-terminal of the delta endotoxin
CC   crystal protein, the synthetic sequence was used as substitution of
CC   the first 138 codons has previously been found to enhance expression
CC   of the protein in plant cells. It can be used in a method of particle
CC   mediated transformation which is effective with various tissue types
CC   of poplar, incl. protoplasts, internode and petiole segments and, more
CC   efficiently with nodule cultures. Transgenic trees produced express
CC   sufficient levels of toxin to provide toxicity to feeding insects.
SQ   Sequence   4583 BP;    1279 A;    1061 C;    1003 G;    1240 T;

Initial Score    =    109  Optimized Score  =    236  Significance =  5.87
Residue Identity =    26%  Matches          =    225  Mismatches   =   219
Gaps             =     56  Conservative Substitutions              =     0

            X         10        20        30        40        50
            CARGAYTG-YWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAAR--ATNMGNG-
              |   ||| |||   | ||  ||| ||     || |   || ||   || | |  |  | |
   TCGGTCCTCCGATCGTTGTCAGAAGTAAG--TTGGCCGCAGTGTTATCAC-TCATGGTTATGGCAGCACTGC
     3940    X 3950      3960      3970      3980      3990      4000

     60        70              80        90        100       110
   --ARYTNWSNGAYTAYY-TNY--TNC--ARGA--YTAYCCNGTNACNGTNGCNWSNAAYY-TNCARGAY---
     |||| ||  ||| || | |  | |  ||||  || | | || ||  | |  || | ||   |||| |
   ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGAC--TGGTGAGTACTCAACCAAGTCATT
      4010      4020      4030      4040      4050      4060      4070

       120       130       140       150       160       170       180
   ----GARGARY-TNTGYGGNGGNY-TNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACN
       ||| ||| | |||||  |  |   |  ||  || |  |     ||||  ||||   ||  |   | |
   CTGAGAATAGTGTATGCGGCGACCGAGTTGC-TCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACA--
```

```
     4080       4090       4100       4110       4120       4130       4140

      190        200        210        220        230        240        250
GTNGCNG-GNWSNAARA-TGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAART--G
 | || |   |  |||| |||     |  || ||||  ||   | | | |    |  ||   |  |  |
-TAGCAGAACTTTAAAAGTGC--TCATCATTGGAAAACGT-TCTTCGGGGGCGAAAACTCTCAAGGATCTTAC
     4150       4160       4170       4180       4190       4200       4210

      260        270        280        290        300        310        320
YGCNTTYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAY-ATNWSNMGNYTNYTNCARGARACNW
|||  || ||   ||     | |||   |   |||| ||  | | | || |  || |  ||   |    |
CGCTGTTGAG-ATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG-CATCTTTTACTTTCAC
      4220       4230       4240       4250       4260       4270       4280

      330        340        350        360        370        380        390
SNGARCARYTNGTNGCNYTNAARCCNTGGATNACNMGNCARAAYTTY----WSNMGNTGYYTNGARYTNCAR
| | |   || |  |  |  ||||   |||      || ||||||  |    |  ||    |  | ||  ||
CAGCGTTTCTGGGTGAGCAAAAAC--AGGA-----AGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACAC
      4290       4300             4310       4320       4330       4340

        400        410        420        430        440        450
TGYCA-RCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNY--TNGARGCNACNGCNCCNACN
 |  | |    ||||    || || ||      |    |     |  | |  |   ||  |  |   |
GGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGA
 4350       4360       4370       4380       4390       4400       4410

460    X
  GCNCCN
  ||
  GCGGATACATATTTGA
4420  X   4430
```

6. US-08-162-407-6 (1-465)
   N90183 DNA sequence of plant vector pAMVBTS

```
ID   N90183 standard; DNA; 4583 BP.
AC   N90183;
DT   1-NOV-1989  (first entry)
DE   DNA sequence of plant vector pAMVBTS
KW   Chimeric gene; transgenic plants; Bacillus thuringiensis;
KW   delta-endotoxin; promoter; translational enhancer; polyadenylation
KW   sequence; tobacco plant; cotton plant; Lepidoptera; Diptera; Coleoptera.
KW   vector; pAMVBTS.
PN   WO8904868-A.
PD   01-JUN-1989.
PF   17-NOV-1988; U04107.
PR   19-NOV-1987; US-123054.
PA   (AGRA) Agracetus.
PI   Barton KA, Umbeck PF;
DR   WPI; 89-178387/24.
PT   Chimeric gene construct for producing transgenic plants - contains
PT   delta-endotoxin of Bacillus thuringiensis for toxicity to, eg
PT   lepidoptera sp.
PS   Disclosure; fig. 4a-b; 55pp; English.
CC   The sequence is of vector pAMVBTS (ATCC 53637), comprising ampicillin
CC   resistance plasmid replicon derived from pMT21, and a chimeric gene which
CC   consists of, in order from the 5' end, a DNA fragment corresp. to
CC   cauliflower mosaic virus 35S transcription promoter, DNA leader fragment
CC   corresp. to alfalfa mosaic virus coat protein mRNA 5' noncoding region, a
```



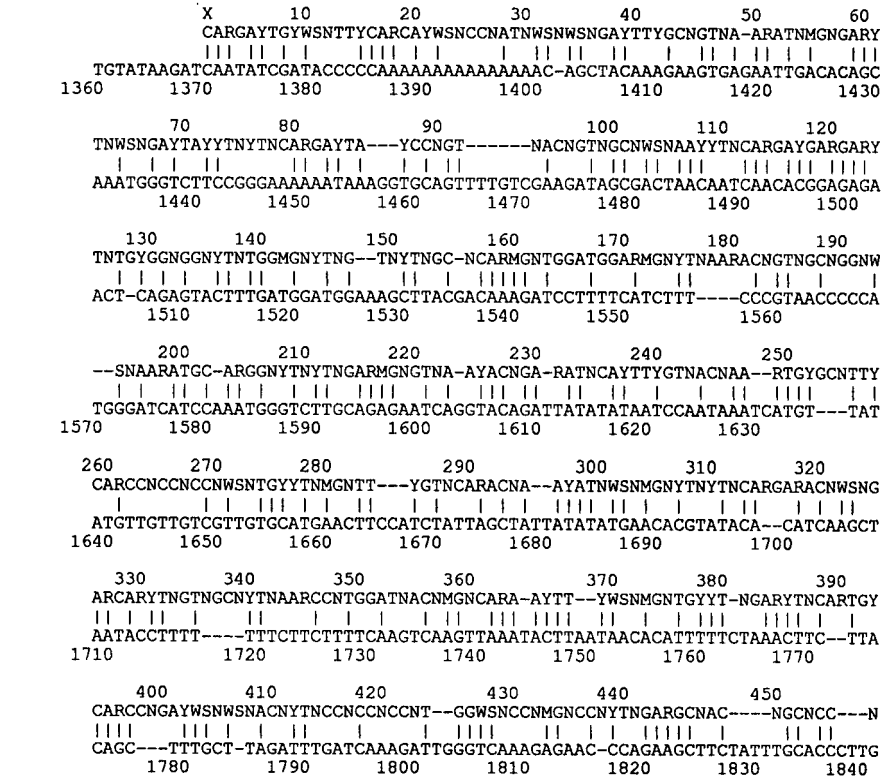
```
CC   DNA fragment corresp. to N-terminal of B. thuringiensis delta-endotoxin,
CC   and a DNA fragment corresp. to polyadenylation region of nopaline
CC   synthase. The vector is used to transform plants, esp. tobacco and
CC   cotton, which then produce the toxin which is active against
CC   Lepidopteran, Dipteran and Coleopteran pests. This trait is then
CC   transmitted by Mendelian inheritance.
SQ   Sequence   4583 BP;   1316 A;   991 C;   977 G;   1299 T;

Initial Score    =    109  Optimized Score  =    236  Significance =  5.87
Residue Identity =    26%  Matches          =    225  Mismatches   =   219
Gaps             =     56  Conservative Substitutions               =     0

             X        10        20        30        40        50
             CARGAYTG-YWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAAR--ATNMGNG-
               |   ||| |||   | ||  ||| ||    || |   || ||   || | |  |  | |
   TCGGTCCTCCGATCGTTGTCAGAAGTAAG--TTGGCCGCAGTGTTATCAC-TCATGGTTATGGCAGCACTGC
     3940    X 3950      3960      3970      3980      3990      4000

     60        70             80          90        100       110
   --ARYTNWSNGAYTAYY-TNY--TNC--ARGA--YTAYCCNGTNACNGTNGCNWSNAAYY-TNCARGAY---
     |||| ||  ||| || | |  | |  ||||  || | | || ||  | |  || | ||   |||| |
   ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGAC--TGGTGAGTACTCAACCAAGTCATT
      4010      4020      4030      4040      4050      4060      4070

      120       130       140       150       160       170       180
   ----GARGARY-TNTGYGGNGGNY-TNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACN
       ||| ||| | ||||| |  |  | ||  || |  |    ||||   ||||   ||  |   | |
   CTGAGAATAGTGTATGCGGCGACCGAGTTGC-TCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACA--
      4080      4090      4100      4110      4120      4130      4140

      190       200       210       220       230       240       250
   GTNGCNG-GNWSNAARA-TGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAART--G
    | || |   |  |||| |||     |  || ||||  ||   | | | |    |  ||   |  |  |
   -TAGCAGAACTTTAAAAGTGC--TCATCATTGGAAAACGT-TCTTCGGGGGCGAAAACTCTCAAGGATCTTAC
      4150      4160      4170      4180      4190      4200      4210

      260       270       280       290       300       310       320
   YGCNTTYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARACNAAY-ATNWSNMGNYTNYTNCARGARACNW
   |||  || ||   ||     | |||   |   |||| ||  | | | || |  || |  ||   |    |
   CGCTGTTGAG-ATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG-CATCTTTTACTTTCAC
      4220      4230      4240      4250      4260      4270      4280

      330       340       350       360             370       380       390
   SNGARCARYTNGTNGCNYTNAARCCNTGGATNACNMGNCARAAYTTY----WSNMGNTGYYTNGARYTNCAR
   | | |   || |  |  |  ||||   |||      || ||||||  |    |  ||    |  | ||  ||
   CAGCGTTTCTGGGTGAGCAAAAAC--AGGA-----AGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACAC
      4290      4300             4310      4320      4330      4340

        400       410       420       430       440       450
   TGYCA-RCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNY--TNGARGCNACNGCNCCNACN
    |  | |    ||||    || || ||      |    |     |  | |  |   ||  |  |   |
   GGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGA
    4350      4360      4370      4380      4390      4400      4410

460    X
  GCNCCN
  ||
  GCGGATACATATTTGA
4420  X   4430
```

7. US-08-162-407-6 (1-465)
Q63241 Crucifer 1-aminocyclopropane-1-carboxylate (ACC) s

ID Q63241 standard; DNA; 5613 BP.
AC Q63241;
DT 14-DEC-1994 (first entry)
DE Crucifer 1-aminocyclopropane-1-carboxylate (ACC) synthase.
KW Crucifer 1-aminocyclopropane-1-carboxylate (ACC) synthase;
KW Arabidopsis thaliana; transgenic plant; ethylene; ripening;
KW maturation; senescence; inhibition; ss.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT GC_signal 340..345
FT /*tag= a
FT CAAT_signal 709..715
FT /*tag= b
FT /note= "Putative CAAT signal."
FT CAAT_signal 723..729
FT /*tag= c
FT /note= "Putative CAAT signal."
FT CAAT_signal 993..996
FT /*tag= d
FT /note= "Putative CAAT signal."
FT CAAT_signal 1151..1154
FT /*tag= e
FT /note= "Putative CAAT signal."
FT TATA_signal 1310..1315
FT /*tag= f
FT /note= "Putative TATA signal."
FT exon 1432..1603
FT /*tag= g
FT /label= Exon 1.
FT intron 1604..1777
FT /*tag= h
FT /label= Intron 1.
FT exon 1778..1909
FT /*tag= i
FT /label= Exon 2.
FT intron 1910..2161
FT /*tag= j
FT /label= Intron 2.
FT exon 2162..2322
FT /*tag= k
FT /label= Exon 3.
FT intron 2323..2595
FT /*tag= l
FT /label= Intron 3.
FT exon 2596..3621
FT /*tag= m
FT /label= Exon 4
FT polyA_signal 3926..2931
FT /*tag= n
FT /note= "Putative polyadenylation signal."
PN WO9408449-A.
PD 28-APR-1994.
PF 13-OCT-1993; U09816.
PR 15-OCT-1992; US-962481.
PA (GEHO) GEN HOSPITAL CORP.
PA (UYGE-) RIJKSUNIV GENT.

PI Goodman H, Van DER STRAETEN D, Van MONTAGUE M;
DR WPI; 94-150818/18.
DR P-PSDB; R53114.
PT DNA encoding crucifer ACC synthase - used for prodn. of
PT recombinant polypeptide(s) or regulating ethylene-inducible
PT events in a plant.
PS Claim 1; Figure 1A; 54pp; English.
CC The crucifer 1-aminocyclopropane-1-carboxylate (ACC) synthase coding
CC sequence can be used to produce transgenic plants in which ethylene
CC inducible events such as fruit ripening, fruit maturation or
CC senescence are inhibited.
SQ Sequence 5613 BP; 1794 A; 959 C; 960 G; 1900 T;

Initial Score = 108 Optimized Score = 236 Significance = 5.81
Residue Identity = 28% Matches = 228 Mismatches = 215
Gaps = 64 Conservative Substitutions = 0



```
    ACNG----CNCCN
    |  |    |  |
    AAGGTATTCATCAGTTTAGCGAC
         1850   X  1860
```

8. US-08-162-407-6 (1-465)
   N30044 Sequence encoding bovine leukocyte interferon (IFN

```
 ID   N30044 standard; DNA; 1162 BP.
 AC   N30044;
 DT   25-MAY-1992  (first entry)
 DE   Sequence encoding bovine leukocyte interferon (IFN) -beta-2.
 KW   Antiviral; cancer; therapy; immunosuppression; tumour; ss.
 OS   Bos taurus.
 FH   Key             Location/Qualifiers
 FT   sig_peptide     320..382
 FT   /*tag=  a
 FT   mat_peptide     383..880
 FT   /*tag=  b
 PN   EP--88622-A.
 PD   14-SEP-1983.
 PF   08-MAR-1983; 038114.
 PR   08-MAR-1982; US-355298.
 PR   01-NOV-1982; US-438128.
 PA   (GETH ) GENENTECH INC.
 PI   Capon DJ, Goeddel DV;
 DR   WPI; 83-766437/38.
 DR   P-PSDB; P30078.
 PT   Polypeptide(s) with amino acid sequence of an animal interferon -
 PT   useful as antiviral, anticancer and immuno-modulator agents
 PS   Disclosure; Fig 9b; 62pp; English.
 CC   A bovine genomic DNA library was screened for IFN genes.  Four
 CC   bovine leukocyte IFN genes were identified (see N30039, N30040,
 CC   N30041, N30042).  Whereas the BoIFN-alpha-2 and -3 genes encode
 CC   peptides with only minor apparent differences to BoIFN-alpha-1, the
 CC   BoIFN-alpha-4 protein is as distinct from the other bovine peptides
 CC   as are any two bovine and human leukocyte interferons.  Three
 CC   distinct bovine fibroblast IFN genes were also identified (see
 CC   N30043, N30044, N30045).  The bovine proteins are quite distinct
 CC   from one another but show an even greater difference with the human
 CC   peptide.
 SQ   Sequence   1162 BP;    371 A;    257 C;    234 G;    300 T;

Initial Score    =    100  Optimized Score  =    251  Significance =  5.29
Residue Identity =    31%  Matches          =    242  Mismatches   =   187
Gaps             =     64  Conservative Substitutions              =     0

          10  X     20        30        40        50        60
    RGAYTGYWSNTTYCARCAYWSNCCNATNWSNWSNGAYTTYGCNGTNAARATNMGNGARYTNW--SNGAY---
              |||  ||||||   | ||    ||  | ||| |    || || ||  |||| |  | |||
              TTTTAGCATTTAGCAATTCACTGAAAATTTAC-AAAAACATTAG-AAATTCTCCCAGACTGT
              X       10        20        30        40        50        60

    70             80        90        100       110       120       130
    -TAYYTNYTNC-----ARGAY-TAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYTNT-GYGGNG
     |||||  || |     || || |||    | |     || || | ||   ||||  |||||||  | | ||
    ATATCTTTTTCCCCTTAATACATATAAAATCA-AAAAGCAAGGAGCTAAAAAGAAAAAGAGTTTTAGAGGTA
            70        80        90        100       110       120       130
```



```
          140       150       160       170       180       190       200
    GNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGA---TGGARMGNYTNAARACNGTNGCNGGNWSNAA-RATG
      |  |  |  |  |     |  | ||   |||   |||||||    |||   |  |    || || ||||
    ACTAATCAACACAGG----AGAACTAAAAAGGAAACTGGAAAGTGGGAAATTCCTCTCCAATAGAAAGAATG
          140       150       160       170       180       190

           210       220       230       240       250       260       270
    CARGGNY-TNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNACNAARTGYGCNTTYCARCCNCCNCCNWS
      |||| | | || | |   || | | |  | |  ||    |  |    || | |  | ||| |    || ||
    GAGGGCCATGCT-GTATAAGT-AGCCC--ACACTCAAGAAGGAAGGCCATTCACTCTGCAAAC----CCTTG
200       210       220       230       240       250           260

          280       290       300           310       320       330
    NTGYYTNMGNTTY-GTNCARACNAAYATNW---SNMGNY-TNYTNCARGARACNWS-NGARCARYTNGTNGC
      | || || ||| |  |  || | ||  |   | || | | |  | |    | ||  || | ||  ||  |
    AAGACTCAGCTTCAGCACCTACTAGCAGAACAGGCAGCCCTGTGCCTGATTTCATCATGACCCACCGGTGCC
        270       280       290       300       310       320       330

    340       350       360       370           380           390       400
    -NYTNAARCCNTGGATNACNMGNCARAAYTTYWS-----NMGNTGYYT---NGARYTNCARTGYCARCCNGA
      ||  ||   ||| |  | |  | |  ||||||      || | |||    |||||  |||  |    | ||
    TCCTCCAG--ATGGTTCTC-CTGCTGTGTTTCTCCACCACAGCTCTTTCCAGGAGCTACAGCTTGCTTC-GA
      340       350       360       370       380       390       400

          410       420           430       440           450       460    X
    YWSNWSNACNYTNCCNCCNCCNTG-GW---SNCCNMGNCCNYTNGARGC---NACNGCNCCNACNGCNCCN
    |||  | ||     |    || || ||   | |  |  |  || |||||    ||  |  | ||  | |
    TTC--CAACAAAGGCGGAGCCTTGAGTTATGTCAGAAACTCCT-GAGGCAGTTAC--CTTCAACTCCTCAAC
         410       420       430       440       450           460       470

    ATTGCCTCG
```

9. US-08-162-407-6 (1-465)
   Q14235 A.faecalis penicillin acylase coding sequence.

```
 ID   Q14235 standard; DNA; 2451 BP.
 AC   Q14235;
 DT   10-JAN-1992  (first entry)
 DE   A.faecalis penicillin acylase coding sequence.
 KW   antibiotic; pac gene; beta-lactam acylase; EC.3.4.1.11;
 KW   ss.
 OS   Alcaligenes faecalis ATCC 19018.
 FH   Key             Location/Qualifiers
 FT   sig_peptide     1..78
 FT   /*tag=  a
 FT   /evidence= experimental
 FT   mat_peptide     79..708
 FT   /*tag=  b
 FT   /label= alpha-subunit
 FT   mat_peptide     709..2448
 FT   /*tag=  c
 FT   /label= beta-subunit
 PN   EP-453047-A.
 PD   23-OCT-1991.
 PF   18-APR-1991; 200929.
 PR   18-APR-1990; EP-200962.
 PR   20-DEC-1990; EP-203483.
 PR   18-APR-1991; EP-200929.
```

```
PA   (KONN ) GIST-BROCADES NV.
PI   Quax WJ;
DR   WPI; 91-312307/43.
DR   P-PSDB; R14444.
PT   Penicillin G acylase gene - isolated from Alcaligenes faecalis,
PT   used to produce recombinant enzyme for prodn. of penicillin(s)
PT   and cephalosporin(s)
PS   Claim 1; Page 8; 23pp; English.
CC   An A.faecalis ATCC 19018 genomic DNA library was constructed from
CC   4-7kb fragments of a partial Sau3A digest. A probe was designed from
CC   the N-terminal sequence of the beta-subunit which had been isolated
CC   from A.faecalis cells. Two positive clones were isolated. Acylase
CC   activity was assigned to a subclone of one of the positive clones.
CC   The subclone was sequenced and the ORF is given here.
SQ   Sequence   2451 BP;    534 A;    680 C;    739 G;    498 T;

Initial Score     =    100  Optimized Score  =    246  Significance =  5.29
Residue Identity  =    30%  Matches         =    246  Mismatches   =   188
Gaps              =     71  Conservative Substitutions              =     0

            X             10        20        30        40        50
            CARGA-YTGY------WSNTT-YCARCAYWSNCCNATNWSNWSNGAYTTYG--CNGTNAARA
            ||| | ||||      ||  |  ||||| || |     || || ||| |||  |      ||
   AAAGCTGGCGCAACAGCTGCCGGCCGACGATAGCAGCA-AGGCGGCGCTG-AC-GATGTTGCTCGCCTGGGA
         1710      1720      1730      1740      1750      1760

          60        70        80        90       100       110
   T--NMGNGARYTNWSNGAYTAYYTNYTNCARGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAY-----
   |   |  ||||  || ||| |       ||| ||   || |  || |  ||  |   ||| ||||||
   TGGAATGGAAC--AG-GATCAGGGAGGGCAAAAT--GCCGG--ACCGGCGCGGGGTGCTCTTCAAGACCTGGC
1770      1780          1790      1800          1810      1820      1830

     120       130       140       150       160       170       180
   --GARGARYTNTGY--GGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNY-TNAARACNG
     |||||| | | |  | || || || | |  | ||  | || | ||   | || | ||  | |  || | |
   TGGAAGAAATGTACAAGCAGGTCT-TG-ATGCCGGTGGT-GC-CTGAATCGCATCGCGCCATGTATAGCCAG
     1840      1850      1860      1870      1880      1890      1900

      190       200       210       220       230       240       250
   TNGCNGGNW---SNAARATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCA-YTTYGTNACNAARTGY
     | || |   | | |  |||||| |  |     ||  |  ||    || |  | | | |  || ||| ||
   --ACTGGTTTTGCCACGCAGCAAGG-TCCCAACCCCGGTTCCAT---CAACTTGAGCATGGGCACCAAG-GT
         1910      1920      1930      1940          1950      1960

        260       270       280       290       300       310
   GCNTTYCARCCNCCNCCNWSNTGYYTNMGNTTYGTNCARA--CNAA-YATNWSNM-GNYTNYTNCARGAR-A
    | ||   ||     || || ||||  |    | | | ||  | || | | || | | || |     || |
   -C-TTGTTGC--GTGCCTTGGTGCTGGAAGCCCATCCCGATCCCAAGCGTGTGAATGTCTTTGGTGAGCGTT
    1970        1980      1990      2000      2010      2020      2030

  320       330       340       350       360       370       380       390
   CNWSNGARCARYTNGTNGCNYTNAARCCNTGGATNACNMGNCARAAYTTYWSNMGNTGYYTNGARYTNCART
   | ||  || || |  | || |   |||  || |  |  || |||  |  ||  |   |||  ||||
   CGTCTCAGGAAATCAT-GCAC-ACAGCTTTGCAAAATGCG-CAGGCCCGCTTGA---GCCAGGAGCAGGGCG
    2040      2050        2060      2070      2080        2090      2100

         400       410       420           430           440       450
   GYCARCCNGAYWSNWSNACNYTNCC-NCCNCCNTG----GWSNCCNMGNC---CNYTNGA-RGCNACNGCNC
    ||||   |   | || ||  | ||  || || ||    ||  |  || |     ||  | ||  ||  | |
   CTCAGATGGCGCGCTGGACCATGCCGACCTCCGTGCATCGTTTCAGCGACAAGAACTTCACGGGAACCCCGC
```



```
        2110      2120      2130      2140      2150      2160      2170

     460    X
   CNACNGCNCCN
     ||    ||
   AGACGATGCCTGGCAATACCT
     2180   X  2190
```

10. US-08-162-407-6 (1-465)
    Q03661 Maize C3 sequence encoding acetohydroxy acid synth

```
ID   Q03661 standard; DNA; 2546 BP.
AC   Q03661;
DT   08-AUG-1990  (first entry)
DE   Maize C3 sequence encoding acetohydroxy acid synthetase (AHAS) gene.
KW   acetohyroxy acid synthetase; AHAS; sulphonylurea;
KW   herbicide; ds.
OS   Zea sp.
FH   Key             Location/Qualifiers
FT   CDS             256..2169
FT   /*tag=  a
PN   EP-360750-A.
PD   28-MAR-1990.
PF   13-SEP-1989; 810685.
PR   22-SEP-1988; US-248000.
PR   06-JAN-1989; US-294530.
PR   08-JUN-1989; US-363663.
PA   (CIBA) Ciba Geigy AG.
PI   Montoya A, Jen G, Harms C, Carswell G, Armour S, Volrath S;
DR   WPI; 90-093264/13.
DR   P-PSDB; R05589.
PT   Plant cell tolerating herbicidal aceto:hydroxy acid synthase inhibitor -
PT   producing increased levels of enzyme, obtd. by selection
PT   processes or genetic engineering.
PS   Disclosure;  64pp; English.
CC   Using DNA confering tolerance to the effects of AHAS inhibitors such as
CC   sulphonurea herbicide allows for the control of weeds using these
CC   compounds. The tolerance may also be used as a selectable marker in
CC   somatic hybrids and transgenic plants.
SQ   Sequence   2546 BP;    544 A;    717 C;    680 G;    605 T;

Initial Score     =    100  Optimized Score  =    203  Significance =  5.29
Residue Identity  =    27%  Matches         =    195  Mismatches   =   179
Gaps              =     47  Conservative Substitutions              =     0

       60    X   70        80        90       100       110       120
   MGNGARYTNWSNGAYTAYYTNYTNCA-RGAYTAYCCNGTNACNGTNGCNWSNAAYYTNCARGAYGARGARYT
             | |   |||| || || | ||| |   ||    ||    || ||| | | ||  ||| |||
            CCGTAAATCTCTTCCACGCACTCTGTCGT----GT----ACCAACGTGC-AGTGGAAACGCT
            X         10        20        30        40        50

     130       140       150       160       170       180       190
   NTGYGGNGGNYTNTGGMGNYTNGTNYTNGCNCARMGNTGGATGGARMGNYTNAARACNGTNGCNGGNWSNAA
      ||     || || |  |  || |     | |    | |  | ||  ||   ||| ||  | |  ||  |
   -CACGTACCTTTGTGTA-TTATGTACGGATTCGG----GCAACGGACATTT--CGAC-GT--CGGTTTGCGA
        60        70        80        90       100       110

   200       210       220       230       240       250       260
   RATGCARGGNYTNYTNGARMGNGTNAAYACNGARATNCAYTTYGTNA----CNAART-GY--GCNTTYCARC
```

```
  | | | |   |  |  ||||     | | | |||    | |  |  |    |   || ||   | |||| ||
  G-TCCCATTCCCATCTGAACCACACATCTCTGAACAAAAGTAGGGGAGGCGCCCGGTAGCCCCCTTTCCCAC
      120       130       140       150       160       170       180

        270       280       290       300       310          320
  -CNCCNCCNWSNTGYYTNMGNT-TYGTNCARACNAAYATNWSNMGNYTNYTNCARG--ARACNWSNG-ARCA
     ||  |  | ||||  || |  |   |   | ||    || || |  || | ||  || |  | | |||
  AATCCCACTCCGTGCC--AGGTGCCACCCTCCCCAAGCCCTCGCGCCGCCTCCGAGACAGCCGCCCGCAACC
    190       200          210       220       230       240       250

330       340       350       360          370       380       390
  RYTNGTNGCNYTNAARCCNTGGATNACNMG-NCA---RAAYTTYWSNMGNTGY-YTNGARYTNCARTGYC-A
  ||      | |    |||  ||    | ||  ||   |  |  ||  || | | |   |||    | ||| |
  ATGGCCACCGCCGCCACCGCGGCGGCCGCGCTCACCGGCGCCACTACCGCTACGCCCAAGTCGAGGCGCCGA
    260       270       280       290       300       310       320

    400       410       420       430       440       450       460    X
  RCCNGAYWSNWSNACNYTNCCNCCNCCNTGGWSNCCNMGNCCNYTNGARGCNACNGCNCCNAC-NGCNCCN
  |||   ||| ||  | |  ||  |  |    || || || || |  |  ||  | ||     |  || ||
  GCCCACCACTTGGCCAC--CCGGC-GCGCCCTCGCCGCGCCCATCAGGTGC-TCAGCGTTGTCACGCGCCAC
  330       340          350       360       370       380       390   X

  GCCGACGGC
    400
```

maryh@stic **stdin**

NeWSprinter20

Thu Apr 06 10:28:18 1995

NeWSprint 2.5 Rev B
Openwin library 3
NeWSprint interpreter 210.0

NeWSprint 2.5